Author HALDANE
Class Mark S/ QC 171
Book No. 46522

UNIVERSITY
OF NOTTINGHAM
LIBRARY
Science Library

Mining Department.

Presented by

....

....

Catalogue No.[redacted].....

Shelf No................

61 0105573 9

UNIVERSITY OF NOTTINGHAM
WITHDRAWN
FROM THE LIBRARY

Students and External Readers DATE DUE FOR RETURN		Staff & Research Students DATE OF ISSUE
		26. MAR. 1969

Any book which you borrow remains your responsibility until the loan slip is cancelled

UNIVERSITY OF NOTTINGHAM
WITHDRAWN
FROM THE LIBRARY

GASES AND LIQUIDS

All rights reserved

GASES AND LIQUIDS

A CONTRIBUTION TO MOLECULAR PHYSICS

BY

J. S. HALDANE, C.H., M.D., F.R.S.

M.A., HON. LL.D. (BIRM. AND EDIN.), HON. D.SC. (LEEDS)

Fellow of New College, Oxford; Director of the Mining Research Laboratory, and Hon. Professor, University of Birmingham

OLIVER AND BOYD
EDINBURGH: TWEEDDALE COURT
LONDON: 33 PATERNOSTER ROW, E.C.4
1928

UNIVERSITY LIBRARY
12
JAN
1956
NOTTINGHAM

c

PREFACE

ALTHOUGH the reasoning brought forward in the present book must stand or fall on its own merits, some explanation is perhaps called for as to its origin, and how such a book came to be written by one who is not primarily a physicist or physical chemist, but a physiologist.

The book originated in attempts, many years ago, to make use in physiology of the current conceptions, mainly derived from van't Hoff, of osmotic pressure. It was evident that, for instance, the kidneys are constantly engaged in regulating, among other things, what is called the "osmotic pressure" of the blood and tissues, just as the lungs are engaged in regulating the partial pressures of carbon dioxide and oxygen. But if osmotic pressure is a pressure varying directly (in a quite unintelligible manner) as the concentration of total solute molecules present in the blood, it is impossible to bring the phenomena of excretion of water into line with the regulated excretion of various individual solutes or with the regulation of output of carbon dioxide and intake of oxygen by the lungs.

These considerations led me to examine carefully van't Hoff's conception of osmotic pressure and the other phenomena related to it, and to compare the theory with the experimental data. The extraordinarily accurate and complete series of measurements of osmotic pressure by Lord Berkeley and his associates at Oxford afforded a solid basis for criticism and reconstruction; while the older data of Raoult, Pfeffer, de Vries, Arrhenius, Morse, and many other observers, were equally useful in other directions. The new conception to which I was thus led of osmotic pressure and its related phenomena seemed to solve the physiological difficulties, and at the same time to furnish a physical explanation which harmonised with the dynamical theory of gases as well as with the experimental data. I gave a short account of this conception in a paper published in the *Biochemical Journal*, 1918, and also referred in my book on *Respiration*, published in 1922, to its bearings on physiology.

Chapter IV, which deals with osmotic pressure in liquids, was thus the starting-point of the book in my own mind, while Chapters I, III, and V group themselves round this Chapter. I soon found, however, that the theory I had adopted involved modifications of many existing ideas as to the connections between the gaseous, liquid, and solid states. Hence came the reinterpretation in Chapter VI of van der Waals's famous equation, and the application of this reinterpretation to the phenomena of liquefaction, of critical temperatures, and of the internal pressures of liquids. Waterston's fundamental work on the dynamical theory of gases and heat was known to me through the late Lord Rayleigh's discovery and publication of his paper of 1845; and this paper and a reference in Landolt and Börnstein's *Physikalisch-Chemische Tabellen* led me to examine Waterston's papers on the physics of liquids, including a remarkable unpublished paper in the Archives of the Royal Society on the relation of absolute temperature to the molecular concentrations of saturated vapours. The latter papers seemed to me to afford a clue, no less important than that discovered by van der Waals, to the connection between the gaseous and liquid states and the explanation of various common properties of liquids. The discussions in Chapter VII deal with this subject.

Chapters II and VIII deal from the standpoint of molecular physics with the principles on which the action and thermal efficiency of heat-engines depend. Thermodynamical principles are of the greatest importance in physiology. Apart from this my interest in scientific questions connected with the utilisation of coal had forced on me the question why it is that in spite of the enormous amount of first-rate research which has been devoted to the development of heat-engines, the efficiency with which they convert heat energy into mechanical energy remains so low; and why, in particular, the efficiency of steam-engines is specially low. The knowledge that Waterston, whose *Collected Scientific Papers*, edited by myself, are appearing simultaneously with this book, was in disagreement with Kelvin and other leading physicists on the subject of thermodynamics, led me to examine carefully the basis of existing ideas on the subject, and particularly the reasoning of Carnot as to reversible heat-engines. The discussions in Chapters II and VIII represent the substance of this examination, which resulted in a breakaway from certain existing academic ideas on the subject of heat-engines, and an attempt to come to closer quarters with the generalisation commonly known as the second law of thermodynamics. It seems to me that in physiology it is just as necessary to get to close quarters in this direction as in the direction of the phenomena connected with osmotic pressure and the theory of liquids. The great practical importance of applying correct scientific principles in the design and use of engines for obtaining power from heat is also evident enough. The discussion of the steam-engine in Chapter VIII indicates that in the near future small steam-engines of much higher thermal efficiency and greater power are likely to be available for many purposes to which only internal combustion-engines or steam-engines of antiquated design have hitherto been applied.

It was my original intention to extend the scope of the book to solids; but I found that the attempt to do so, except in so far as seemed to be necessary in connection with gases and liquids, was taking me too far from the apparently simple Newtonian principles which had so far been applied. The book may be regarded as an endeavour to develop further Waterston's conception of what he called "Molecular Physics," and hence belongs rather to the nineteenth than to the twentieth century. However clearly we may see that the ideas of the twentieth century are leading us towards a new outlook in physical science, we have still to come to an understanding in detail with the ideas inherited from the nineteenth century. I hope the book may be useful in bringing us nearer to such an understanding. All the new conclusions brought forward seem to make for simplicity; and if I have criticised adversely some of the reasoning of distinguished men of science, including van't Hoff, van der Waals, Carnot, and Kelvin, I think it will be seen that this has not arisen from any lack of appreciation of their great achievements.

To Dr C. G. Douglas, F.R.S., and to other friends, I am indebted for many very useful criticisms of the original drafts of different chapters.

OXFORD, *July* 1928.

CONTENTS

PAGE
PREFACE v

CHAPTER I

THE DYNAMICAL THEORY OF GASES AND HEAT

Introductory Remarks 1
The Gas Laws and their Formulation in Equations . . 2
The Kinetic Theory of Gases and its History . . . 4
Meaning of Temperature and Absolute Temperature . . 6
The Ordinary Gas Equations 10
Defects in the Crude Gas Laws and Corresponding Equations . 12
Improvements in the Gas Equations 14
Van der Waals' Equation 15
Discussion of the "Volumes" of Molecules 16
Compressibility of Molecules 17
Significance of V-*b* in the Revised Gas Equation . . . 18
The Kinetic Theory of Gases—a Dynamical Theory . . 19
Restatement of Gas Laws 20
New Nomenclature of Gas Equations 21
Extended Significance of Gas Equations 23
Relation of Partial Pressure to Molar Concentration at High Pressures 25

CHAPTER II

HEAT-ENGINES AND THE DYNAMICAL THEORY OF HEAT

Carnot's Book on Heat-engines and their Cyclical Action . . 29
Carnot and the Caloric Theory 31
His Demonstration that a Heat-engine working reversibly cannot work more efficiently 32
His Description of a Reversible Heat-engine . . . 33
Graphic Representation of Pressure and Volume in the Cycle of the Engine 34
Waterston on the Graphic Representation and its Wrong Interpretation on the Caloric Theory 37

PAGE
Kelvin's First Interpretation and Thermodynamical Scale of Temperature 38
Later Interpretation by Clausius and Kelvin . . . 39
Waterston and Kelvin's Inferences 40
Fallacies in the Reasoning of Carnot and Kelvin . . . 41
Reversible Heat-engines Impossible 42
Significance of the Fact that Heat is Chaotic Energy . . 48
Thermal Efficiency of an Actual Carnot Engine . . . 53
Its Maximum Theoretical Thermal Efficiency is only a fraction of that of a Reversible Engine 66
Heat-leakage to the Outside 71
Thermal Efficiency, Horse-power, and Effects of Friction and Loss of Heat from the Source 76
Practical Defects of a Carnot Engine 81
Influence of Varying Specific Heat of the Gas . . . 83
A Gas-engine with Internal Heating 85
Internal Combustion-engines 93
Horse-power as distinguished from Thermal Efficiency . . 99
Summary of Argument 100
No need for an Empirical "Second Law of Thermodynamics" . 105
Conclusions of this Chapter bound up with the Mechanical Assumptions of Newtonian Physics 106

CHAPTER III

GASEOUS DIFFUSION AND OSMOSIS

Diffusion of Gases 108
Semi-permeable Septa 109
Definition of Osmotic Pressure of a Gas 110
Equations for Osmotic Pressure and Diffusion Pressure at Low and High Gas Pressures 111
Equation for Osmotic Depression 124
Confusion between Osmotic Pressure and Osmotic Depression . 126
Equations for Isotonic Gas Mixtures 127
Importance of Diffusion Pressures 129

CHAPTER IV

OSMOSIS IN LIQUIDS

Endosmose and Exosmose 130
Semi-permeable Membranes. Graham's and Traube's Discoveries 131
Pfeffer's Experiments on Osmotic Pressure 132

PAGE
De Vries's Experiments on Isotonic Solutions . . . 133
Van't Hoff's Discovery of a Relation between Osmotic Pressure and the Gas Laws 135
His Interpretation of the Relation 138
Defects of this Interpretation 139
Experiments of Morse, and his attempted Reinterpretation . 139
Criticism of van't Hoff's Theory and Statement of True Theory . 140
Experiments of Berkeley and Hartley 145
Experiments of Morse 151
Sources of Error in Morse's Experiments and Calculations . 153
Existing Confusion of Osmotic Pressure with Osmotic Depression 159
Isotonicity of Liquids 161
Discovery by Arrhenius of Osmotic Effects of Electrolytic Dissociation 162
Hydration of Ions 166
Membranes permeable to Anions, but not to Kations . . 167
Influence of Combination with one another of Solvent Molecules . 169
Osmosis and the Posture, Volume, and Spontaneous Movements of Organisms 170
Muscular Contraction 172

CHAPTER V

DEPRESSION OF FREEZING-POINTS AND VAPOUR PRESSURES, AND ELEVATION OF BOILING-POINTS, OF SOLUTIONS

Raoult's Discovery that Depression of Freezing-point is proportional to the ratio of Moles of Solute to Moles of Solvent . 175
The Cryoscopic Constant of each Solvent 176
Parallelism of Depression of Freezing-point and Increase of Osmotic Pressure 178
Van't Hoff's Deduction of Relation between Depression of Freezing-point and Latent Heat of Melting . . . 180
Raoult's Discovery that Depression of the Vapour Pressure of a Solution varies as the ratio of Moles of Solute to Total Moles 184
Relation of Depression of Vapour Pressure to Temperature . 189
Deduction of Raoult's Ratio determining Depression of Vapour Pressure 191
Depression of Vapour Pressure varies parallel to Osmotic Depression, but not to Osmotic Pressure 194
Discussion of an apparent Proof of Arrhenius that Vapour Pressure varies parallel with Osmotic Pressure . . 197
Henry's Law of Solution of Gases 201

PAGE
Elevation of Boiling-point proportional to the ratio of Moles of Solute to Moles of Solvent 203
Deduction of this Law 205
Relation between Elevation of Boiling-point and Latent Heat of Evaporation 207
Mistakes in van't Hoff's Calculation of Elevation of Boiling-point 208
Sources of his Mistakes 214

CHAPTER VI

THE LIQUID AND GASEOUS STATES

The Problems of Liquefaction and the Liquid State . . 216
Hypothesis as to the Location of attractive and repulsive Forces round Molecules 217
Explanation of van der Waals' Equation 224
The Apparent Constants of this Equation are not actually Constant 226
The Latent Heat of Expansion without External Work . . 227
External Pressure and Molecular Chaos 230
Variation of Temperature and van der Waals' Equation . . 231
The Nature of Condensation to the Liquid State . . . 234
Reason for the Existence of a Critical Temperature . . 235
Effects of Condensed Liquid on Gas Pressure . . . 237
Supersaturated Vapours or Gases 240
The Liquid and Solid States 242

CHAPTER VII

THE INTERNAL PRESSURES AND SATURATED VAPOURS OF LIQUIDS

The Diminution of π with increasing Molar Concentration . 244
Molecular Compressibility and Increase of Internal Pressure . 245
The Ramsay-Young Law and Molecular Compressibility . . 249
Van der Waals' Equation and Molecular Compressibility . . 251
Osmotic Pressure and Molecular Compressibility . . . 252
Liquid Expansion and Diminution in Internal Pressure . . 253
Waterston and Liquid Expansion 255
The Expansion of Water 256
Surface Tension 257
The Eötvös-Ramsay-Shields Law 261
Explanation of this Law 262
Internal Pressures of Solutions 265
Adsorption on Solids and Liquids 266
Variation of Internal Pressure in Strong Solutions . . . 268

PAGE
Cause of Constancy of Internal Pressure in not too Strong Solutions 277
Difference between Solutes and Solvents 278
Relation between Temperature and Concentration of Saturated Vapour 278
Waterston's Law 279
Explanation of Waterston's Law 281

CHAPTER VIII

THE STEAM-ENGINE AND SOME OTHER HEAT-ENGINES

An Imaginary Carnot Reversible Steam-engine . . . 286
The Thermal Efficiency of this Imaginary Engine would be the same as that of a Reversible Gas-engine working between the same absolute Temperatures of Isothermal Expansion and Compression 291
The possible Efficiency of an actual Carnot Steam-engine is far less than half that of the Imaginary Reversible Engine . 293
The Logical Error in Carnot's Reasoning about Maximal Efficiency 296
The further Tangle which arose from his assumption that Reversible Heat-engines are possible 298
Actual Reciprocating Steam-engines 300
Thermal Efficiency of Steam-engines 305
Existing Clumsy and Inefficient Steam-engines . . . 310
Efficient Steam-engines and Internal Combustion-engines Compared 311
Locomotive Steam-engines 313
Water-jacket Condensers 315
Double-acting Engines 316
Elasticity of Steam-engines . . . , . . 316
Measurement of Thermal Efficiency 317
Indicator Diagrams 318
Boiler Efficiency 319
Superheating of Steam 319
The Displacement of Steam-engines by Internal Combustion Engines 322
The Future of Steam-engines 324
The Steam Turbine 325
Engines with other Working Substances than Water . . 327
The Future of Solid Coal as a Fuel 328
Muscular Contraction and Carnot's assumption as to Reversible Heat-engines 329
Mechanical Interpretation and Organic Co-ordination . . 330

GASES AND LIQUIDS

CHAPTER I

THE DYNAMICAL THEORY OF GASES AND HEAT

SINCE the atomic theory was established by Dalton and those who followed in his line of reasoning, it has been evident that both in the gaseous and liquid state matter is present in the form of molecules, and that whereas in the gaseous state the molecules are flying free from one another, in the liquid state they are kept in close association by what we are accustomed to call mutual attraction. Gassendi had, indeed, already suggested this as early as 1658. The dynamical or kinetic theory of gases and heat, propounded first about the middle of last century, has enabled us to predict the most striking physical properties of permanent gases at ordinary pressures and temperatures; but we are still without a correspondingly clear insight into the physical properties of imperfect gases, liquids, and solids, or into the causes which limit the transformation of heat energy into mechanical energy.

The theory of van der Waals, based on his famous equation, seemed indeed to point towards a means of connecting the behaviour of gases with that of liquids. When, moreover, van't Hoff[1] made the great discovery that the "osmotic" pressure observed in connection with solutions of sugar and other organic solutes at great molecular dilutions corresponds closely with the pressure which the solutes would exert if they were present in the gaseous form in the same concentrations and at the same temperature, and if Boyle's and Charles's laws held good for them, it became quite evident that the dynamical theory must be capable of extension in some exact form to liquids. Nevertheless the theory of van der Waals has proved to be of only limited service, and there has been as yet no real agreement as to the nature of osmotic pressure, nor, consequently, as to the definite form in which the dynamical theory can be extended to liquids. As a consequence of this obscurity there is much confusion in physical chemistry, as well as in physiology.

---

[1] Van't Hoff, *Zeitschr. f. physik. Chemie*, vol. i., 1887.

Before any discussion of liquids, it is absolutely necessary to discuss certain fundamental points connected with the dynamical theory of gases and the "gas laws" which that theory enabled us to predict. What are generally known as the chief gas laws are three in number. The first is Boyle's or Mariotte's law,[1] that at a given temperature the volume of a given quantity of gas varies inversely as its pressure. The second law is that of Charles, or of Dalton and Gay-Lussac,[2] that the pressure of a given quantity of gas at constant volume increases or diminishes, for every degree of its rise or fall of temperature, by a constant proportion of its pressure at freezing-point or any other convenient temperature.

---

[1] Robert Boyle, *New Experiments, Physico-mechanical, etc., touching the Spring of the Air*, 1662. Mariotte discovered the same law independently in 1676.

[2] This law was first published by Dalton in 1801, and then in 1802 by Gay-Lussac, who, however, mentioned that it had been discovered fifteen years earlier by Prof. Charles of Paris, though not published.

It follows from these two empirically discovered laws that if we represent by a the proportion by which a given quantity of gas at constant volume increases in pressure on rising from 0° to 1° C., and temperature centigrade, positive or negative, by t, pressure by p, and volume by v, we obtain for a given quantity of gas, in so far as these two laws hold good, the equation $\frac{pv}{1 \pm at} = r$, where r is a ratio varying only with the units of pressure and volume employed. The equation may also be written $pv = r(1 \pm at)$. Since the exact value of a is $\frac{1}{273.1}$ this gives us an absolute zero of $-273.1°$ C., since the pressure, according to the equation, would fall to zero at that temperature.

The third of the gas laws is that first formulated tentatively by Avogadro,[1] and gradually established firmly by the work of Gerhardt and other chemists. According to this law equal volumes of different gases at the same temperature and pressure contain equal numbers of molecules, regardless of their weight or size. This law is based partly on the discovery by Gay-Lussac that gases combine with one another in simple proportions by volume, and partly on the same class of facts as justified the atomic theory. Avogadro's law implies that the mass of gas present in a given volume of gas at any given temperature or pressure varies directly as the molecular weight of the gas. If, therefore, we take masses of different gases representing respectively the amount in grammes of the molecular weights of these gases (*i.e.*, gramme-molecules, or moles), they will occupy equal volumes at equal pressures and temperatures, and will contain the same number, n, of moles. The gas equation can therefore be put into the form $pv = nr(1 \pm at)$, and then connects the mass in grammes and temperature centigrade of any gas or mixture of gases with the pressure and volume in whatever units are employed.

---

[1] Avogadro, *Journ. de Phys.*, 1811.

To the three chief gas laws Graham[1] added a fourth: namely, that the rate of diffusion or effusion of a gas varies inversely as the square root of its molecular weight. This also was a purely empirical law, based solely on the results of experiments.

Until nearly the middle of last century the gas laws were simply the unconnected results of empirical observations. No really intelligible theory had been successful in connecting them together, or bringing them into relation with other observed facts relating to gases. This connection, and at the same time an elucidation of the connection between mechanical work done by a gas and fall of its temperature, was first pointed out in a paper submitted to the Royal Society in 1845, by J. J. Waterston, Naval Instructor in the East India Company's College for Cadets, Bombay. The paper, though written in very clear and simple language, and though its reasoning is a foundation stone of physical chemistry, was most unfortunately rejected, as not worthy of publication, by the referees appointed to examine it. The manuscript, in accordance with the statutes, became the property of the Society, and was deposited in the "Archives," where, though available for consultation, it remained unheeded till its discovery long afterwards by the late Lord Rayleigh. It was published, with a very full introduction and valuable notes by him, in 1892.[1] Waterston had also stated the dynamical theory in its main outlines in a physiological book which he published in 1843, but which attracted no attention. He also summarised it in 1845 in a short printed pamphlet, which he likewise sent to the Royal Society, and which is reprinted in his *Collected Papers*. The title of Waterston's paper was, "The Physics of a Medium consisting of free and perfectly Elastic Particles." It is evident from his previous and subsequent writings that the reason why he did not choose a simpler title, such as "The Physics of Gases," was that he thoroughly realised that molecular behaviour cannot be fully described in simple terms of his theory. The "medium" dealt with in his paper was therefore only an ideal one, though representing very closely, as he proceeded to show, the most prominent characters of permanent gases. In actual fact, there is no such thing as a gas of which the molecules are perfectly "free": in other words, no gas is "perfect" in the sense of conforming to the theory exactly, even when allowance is made for the volume of the molecules themselves.

---

[1] Graham, *Transactions of the Edinburgh Royal Society*, vol. xii. 1834. (Read 1831.)

---

[1] Waterston, *Philosophical Transactions of the Royal Society*, A, p. 1, 1892, and page 207 of his *Collected Papers* (Oliver and Boyd, 1928).

From the conclusion, based on a combination of chemical and physical evidence, that gas consists of freely-moving, separate molecules, so that the pressure of the gas is due to chaotic bombardment at all angles and all sorts of velocities by these molecules, it is evident that, other things being equal, the pressure on the walls of the containing vessel must vary in proportion to the frequency of the impacts. Hence, since this frequency must increase in proportion as the volume of the gas is diminished, the pressure must increase similarly. There is thus no difficulty in accounting for Boyle's law; and Daniel Bernoulli had, indeed, so accounted for it in 1738,[1] though his theory, for what then seemed good reasons, had not been accepted. Waterston's paper dealt mainly with the meaning of Charles's, Avogadro's and Graham's laws. He showed that with varying mean velocity of the molecules the pressure must vary in proportion to the combined *vis viva* of the molecules impinging on the containing surface in unit of time, and therefore does not vary directly as the velocity, but as the square of the velocity. If, therefore, as there was even then every reason to believe, rise of temperature is accompanied by an increasing mean velocity of the molecules, each degree of rise, since with the volume constant it produces an equal rise in pressure, must be proportional to increase in the mean square of the velocities of the molecules.

---

[1] Daniel Bernoulli, *Hydrodynamica*, 1738.

In the light of this conclusion, Waterston also concluded that the zero to which the indications of a gas thermometer point must be a condition in which the molecules would be at rest relatively to one another; also that when we are measuring the temperature of a gas by the variations in its pressure or volume, we are really measuring the mean of the squares of the relative velocities of its molecules, since by the law of probabilities the pressure will, when sufficiently large numbers of molecules are concerned, depend simply on this mean square.

Heat is thus chaotic molecular energy of motion, to use modern language; and he inferred from the existing data as to the value of the constant *a*, that "absolute" zero (−461° Fahrenheit or −273.9° centigrade, according to the existing data) is the temperature at which molecules are at rest relatively to one another. In 1848, as is well known, William Thomson (Lord Kelvin) interpreted the zero of gas temperature on the basis of part of Sadi Carnot's reasoning as to heat-engines. Thomson was, of course, quite ignorant of Waterston's paper and, as will be shown in Chapter II., the basis of his interpretation, which has survived to the present time, was unsound.

When a gas expands and does work on a piston or other movable wall of the containing vessel, in virtue of its pressure being greater than that of the atmosphere, this is because the pressure produced by molecular bombardment is greater on the inside. It is thus the collective kinetic energy of the molecules which is transmitted to the piston and engine; and the temperature of the gas, when allowance is made for its specific heat, must, on the dynamical theory of gases, fall in correspondence with the loss of energy. This was pointed out by Waterston, who thus explained, without resort to the unintelligible idea then prevalent that heat becomes "latent," how it is that gases cool themselves as they expand in the cylinder and do work.

By Avogadro's law, equal numbers of molecules or moles of gas occupy the same volume at the same pressure and temperature, whether the molecules are relatively heavy or light, large or small. This remarkable fact can only mean, as Waterston pointed out, that the mean relative *vis viva* (mass multiplied by the square of the velocity) is the same for all the gaseous molecules. The light molecules, such as those of hydrogen, must move faster in inverse proportion to the square root of their molecular weight. He argued that this follows from the laws of impact of enormous numbers of free elastic bodies. By the law of probability the kinetic energy of these bodies must, in a perfect gas, distribute itself evenly on an average, just as the bodies themselves become, by diffusion, evenly distributed. Graham's law of diffusion, that the rate of diffusion of a gas varies inversely as the square root of the molecular weight, was also accounted for in this way.

Waterston's great advance consisted in his demonstration that the *vis viva* of an average gas molecule in a theoretically "perfect" gas varies in proportion to the theoretical absolute temperature, and at the same temperature must be on an average equally distributed among all the molecules, regardless of whether they are heavy or light, large or small. At the time when his paper was written, the confused "caloric" theory was the basis of the accepted mathematical physics of gases, and the general principle of conservation of energy had not been accepted.[1] We have only to substitute energy of translation and rotation for *vis viva* in order to express his theory in more modern language. He had pointed out that the apparent loss of *vis viva* when inelastic substances collide is only apparent. In becoming heated the substances conserve their *vis viva*; but the theory could not by itself give any clear account of latent heat of gaseous expansion against no resistance; and Waterston's assumption that the molecules of his ideal gas were not only perfectly elastic, but also perfectly free, excluded this discussion. His assumption also took no account of radiation, and implied that the specific heat of a perfect gas is the same at all temperatures.

Twelve years after Waterston's paper was presented to the Royal Society, a similar explanation of the two first gas laws was, quite independently of Bernoulli and Waterston, brought forward by Krönig and developed by Clausius. In 1860 the kinetic theory was extended, by similar reasoning to that of Waterston, to the third and fourth gas laws by Clerk Maxwell. Willard Gibbs, Boltzmann, and many other subsequent investigators, developed the theory further. In spite of scientific conservatism the kinetic theory gradually triumphed, and has proved extraordinarily fruitful in many different directions. It is not necessary to follow all of these developments here, as the concern of this book is mainly with the connection between gases and liquids, and with the conditions determining the conversion of heat into mechanical energy.

---

[1] The first clear formulation of this principle was in an almost unnoticed paper by J. R. Mayer in *Liebig's Annalen*, 1842; and Mayer's famous book, applying the conception in physiology, appeared in 1845. In 1847 Helmholtz, another physiologist, independently formulated the same principle on a still wider basis.

Certain deductions from the kinetic theory of gases must now be referred to specially. It is evident that in the light of the kinetic theory the equation for a gas can be considerably simplified. The value of a in the equation quoted above has been found by careful experiment to be $\frac{1}{273.1}$ when t is in centigrade units. Hence, on the kinetic theory 0° C. corresponds to 273.1° on an absolute scale of temperature, and this absolute scale can be denoted by T, so that the equation becomes $pv = nr\frac{T}{273.1}$. But when p is expressed in atmospheres, and v is expressed in litres, the value of pv, when p, v, and n equal one, is about 22.41 at 0° C. Thus one mole of gas at 0° C., or 273.1° of absolute temperature, and atmospheric pressure, occupies about 22.41 litres. This figure has been obtained from numerous exact experiments, including those of Rayleigh. If, therefore, we take P to denote pressure in atmospheres, V to denote volume in litres, and n to denote number of moles of gas present, the equation, according to the theory, becomes $pv = n\frac{22.41\ \mathrm{T}}{273.1}$; and this simplifies to $pv = n\,.08206\ \mathrm{T}$. At ordinary pressures and temperatures, and for permanent gases—*i.e.*, such as cannot be liquefied at any ordinary temperature—this equation holds good almost exactly, whether the gas is pure or a mixture of gases. We need only know the number of moles present, and the absolute temperature, in order to predict the pressure from the volume, or *vice versa.* Similarly, we need only know the pressure and volume in order to predict the number of moles present from the absolute temperature, or *vice versa.* With the volume and moles of gas known we can of course predict the absolute temperature from the pressure at any given temperature or *vice versa.*

This gas equation is also commonly written in the form $\mathrm{PV} = \mathrm{RT}$. In this form PV means the particular value of pv, in any units, when one mole of gas is present.[1] The value of r is then, by Avogadro's law, the same for all gases, provided the units of pressure and volume are the same, and is therefore written R. Its value is .082 when P is expressed in atmospheres, V in litres, and T in centigrade degrees.

When several perfect gases are present at an ordinary pressure, it is evident that each of them must contribute to the value of pv a share which is proportional to the value of n for the gas in question. Thus, if v is constant, part of the value of p is determined by each gas present, and each part is known as the partial pressure of the gas in question. This fact is known as Dalton's law, since Dalton established it by experiment; but, of course, Dalton's law follows from the other laws already referred to, and from the kinetic theory.

---

[1] The real meaning of PV, when the gas equation is put into the form $\mathrm{PV} = \mathrm{RT}$, is very apt to be forgotten. A better form of the equation is perhaps $pv = n\mathrm{RT}$.

The gas equations just discussed hold pretty satisfactorily for permanent gases at or below atmospheric pressure, and at or above ordinary temperatures; but we have now to consider the case of gases which are highly compressed or greatly cooled. As has been well known for long, all gases when highly compressed or greatly cooled cease quite evidently to follow the gas laws in the form in which they are stated above; and the deviations from these laws may be in opposite directions. Thus, on increasing the pressure, the volume may either diminish to a less or to a greater extent than corresponds to Boyle's law. Similarly, on continuously cooling a gas at constant pressure the volume may either diminish to a greater or to a less extent than corresponds to Charles's law. When the volume diminishes out of proportion to Boyle's or Charles's law, the gas may, at a certain pressure or temperature, begin to condense to a liquid or even solidify; but with further increase of pressure or lowering of temperature the diminution in volume always becomes far less than corresponds to Boyle's or Charles's law. A liquid at ordinary temperature, or a solid, is far less compressible than a gas, and diminishes far less in volume as the temperature falls. If, moreover, the gas does not actually condense to a liquid on raising the pressure, it finally begins to diminish in volume less than corresponds to Boyle's law.

These facts have been established by the careful experiments of various observers, and particularly by those of Amagat. As it is also evident from many facts that molecules attract one another, it is agreed by all that the molecular attraction is the cause of a gas diminishing in volume to a greater extent than corresponds to Boyle's or Charles's laws. It has also been generally supposed that as the molecules of gas are brought nearer to one another, they come further and further into the spheres of their mutual attraction, which thus has a gradually increasing effect in compressing them all together. The further discussion of this hypothesis will be postponed to Chapter VI. When liquid is formed, the attraction is so great in relation to the repulsive forces that the molecules, except for a few with exceptional velocities at the surface, can no longer fly free from one another.

On the other hand, the reason why a gas or liquid may diminish in volume less than corresponds to the gas laws is evidently the fact that the molecules themselves are relatively incompressible and incapable of contracting on cooling, so that with increasing diminution in the gross volume of a given mass of gas or liquid a larger proportion of this gross volume is made up of the relatively unalterable "dead" space occupied by the molecules. The free space in which the molecules fly about is thus diminished more than is the gross volume of the gas, and there is a correspondingly greater increase in the pressure, as Dupré[1] and Hirn[2] pointed out.

To allow for the dead space (the "co-volume," as Dupré called it) van der Waals designated this volume by b, so that the gas equation takes the form $P(V - b) = .082T$,[3] or $P(V - b) = RT$, where, of course, the value of b is that for one mole of gas. This form of the equation is wholly consistent with the dynamical theory in its original simple form. To give recognition to the fact that the molecules collide with extra translational kinetic energy or *vis viva* owing to their mutual attractions, Hirn further modified the equation by substituting $p + R$ for P, the extra pressure being denoted by R, and p being the externally transmitted pressure, which for a liquid is nothing more than the vapour pressure. Hirn's equation $(p + R)(V - b) = .082\,T$, represents a great potential step forward, as will appear later, but throws no light on the evidently variable value of R. The equation serves to indicate that P, the internal bombardment pressure in a collection of molecules, is greater, even if only very slightly so, and at great molecular concentrations may be enormously greater, than p the pressure measured externally. The equation $P(V - b) = .082\,T$, still, however, holds good, though

---

[1] Dupré, *Annales de Chemie et de Physique* (4), I. p. 168 (1864).

[2] Hirn, *Théorie mécanique de la chaleur.*

[3] If we allow for the volume of the molecules, the value ·08200 seems to be more nearly correct than ·08206.

P is no longer merely the externally measured pressure, but the internal bombardment pressure. It is in this sense that P will be used in this book.

In 1873 van der Waals published the extremely important discovery[1] that if for the ordinary gas equation $P(V-b) = RT$, we substitute the equation $\left(p + \frac{a}{V^2}\right)(V-b) = RT$ (when p represents external pressure and V volume occupied by one mole), and select suitable values, which vary with different gases, for a and b, the equation enables us to predict with remarkable success the changes in volume which actually occur at a given temperature with varying pressures, provided that the gas is not compressed too far, and does not condense to a liquid. Liquefaction is, however, a phenomenon which van der Waals's conception of the meaning of the equation does not help us to predict, nor does the equation hold at very great concentrations of the gas. The constants a and b in this equation can only be decided on empirically from the results of experiments at different pressures on each variety of gas; and the value which fits each constant depends on the value of the other. It is evident that if we substitute $\frac{a}{V^2}$ for Hirn's R we bring Hirn's equation into line with that of van der Waals.

Van der Waals considered that the a in this equation represents an attractive force which varies inversely as the square of the volume, while of course the b represents the correction for the dead space. In reality, as will appear in Chapter VI., where van der Waals's equation will be discussed, the equation is only capable of giving an approximation, though an extremely useful one, to the changes in pressure and volume which occur in very concentrated gases; but at moderate and low concentrations the equation applies very closely. Van der Waals's interpretation of his equation is, however, not only very improbable, but, as will appear below, would make it impossible to extend the dynamical theory to the phenomena observed in liquids.

---

[1] Van der Waals, *The Continuity of the Liquid and Gaseous States.* English translation published in 1891.

The nature of the dead space b, the question of molecular compressibility, and the variation of intermolecular attraction with distances of the molecules from one another, cannot be dealt with fully until the phenomena of the liquid and solid states are discussed. But meanwhile it is necessary to reach some preliminary conceptions as a basis for interpreting these phenomena.

In the light of the molecular interpretation of matter, it is evident that both the cohesion of solids and the mass-cohesion of liquids imply the existence of powerful attractive "forces" between molecules which are very close to one another. The existence of the gaseous state shows that these forces become practically inappreciable at a very short distance away. On the other hand, the resistance of solids and liquids to compression shows that molecules themselves oppose enormously powerful repulsive forces to compression, and must therefore be comparatively incompressible.

In former times, atoms or molecules were conceived as "solid" bodies which bumped abruptly against one another in their passage through a surrounding "vacuum" or "ether," and from the surfaces of which forces of attraction or repulsion, acting through the vacuum, originated. This conception is quite inconsistent with what is now known of atoms and electrons. The surface of a molecule, if we still think of "surface" in the old sense, can only be a virtual or ideal surface, defined by the points at which compression by attraction and bombardment may be regarded for mere statistical purposes as, on an average, neutralising repulsion round the centre of a molecule.

The *virtual* shapes of different kinds of molecules, as marked out in this sense, are doubtless very various, but must always be more or less rounded. It follows also that when two molecules collide with one another their *virtual* surfaces (if we extend in imagination the conception of virtual volume to individual molecules) coalesce, as if the molecules fused with, or embraced, one another, and formed a continuous virtual molecule with rounded outlines. The virtual spaces between the virtual surfaces of several colliding molecules are thus greatly diminished, or, with favourable conditions of impact, completely abolished momentarily. These virtual spaces must also be regarded as rounded in outline, in correspondence with the rounded outlines of the virtual spaces occupied by the molecules themselves and colliding molecules.

The virtual volume of a molecule, since it is diminished with increase in the ratio of forces of compression to forces of repulsion, must be variable. In other words, a molecule or atom must be regarded as compressible, although on account of the rapidity with which resistance to compression increases, the compressibility of a mass of molecules will be practically inappreciable except under great pressures. With sufficiently great pressures a point must finally be reached where with further increase of pressure the diminution in b becomes quite sensibly proportional to the diminution in $V - b$. The further development of this reasoning will, however, be postponed to Chapter VI.

If the molecules were packed sufficiently close together, but still in motion, then supposing that at a given instant we could obtain a virtual picture showing the distribution of the virtual volumes of the molecules and the "empty" spaces between them, this picture would, on the dynamical theory as now presented, show a largely continuous structure, dotted about with more or less spherical interspaces of various sizes. But these spaces would be on an average proportional in volume to the total number of molecules present and inversely proportional to the pressure; also, at constant pressure, directly proportional in volume to the absolute temperature. The interspaces would, of course, be constantly appearing and disappearing, or altering their individual volumes.

On the theory just outlined the value of the expression $V - b$ in the gas equation, where b signifies the virtual volume of the molecules in one mole for statistical purposes, varies inversely as the pressure on the molecules increases, and directly as the absolute temperature and the number of molecules present. This simplifies enormously the quantitative treatment of the subject.[1] That the simplified treatment is justified by a mass of experimental data will appear in the succeeding chapters. It may be remarked here that the assumption of molecular rigidity and incompressibility is no part of the dynamical theory of actual gases. At any rate, it was no part of Waterston's original conception of actual gases, as appears in the outline which he published in 1843.[2] Subsequent writers have assumed, on what will be shown later to be inadequate grounds, that molecules are rigid and incompressible; but the perfect molecular elasticity which is the basis of the dynamical theory becomes at once intelligible, together with many other observed facts, when the surfaces of molecules are regarded as simply virtual surfaces, the existence of which is only assumed in order to facilitate statistical treatment.

It will now be evident that the expression "kinetic" theory, as applied to the gaseous and other states of matter, is misleading. We must assume that it is not either movement or *vis viva* (kinetic energy) that is conserved among the molecules, but energy. The word "kinetic" refers us back to a time when, although Mayer's first enunciation of the wider conception of energy was in 1842, a year before Waterston's first enunciation of the dynamical theory of gases and heat, this wider conception of energy was still not generally realised. When two molecules collide they must be regarded as gradually losing, and then again regaining, their *vis viva* or kinetic energy with reference to one another. In this process, kinetic can be regarded as converted into potential energy, and afterwards potential into kinetic energy, so that what is conserved is energy. With the disappearance of the unintelligible conception of rigid and incompressible atoms, the expression "dynamical theory" ought to be substituted for "kinetic theory" if we are to avoid confusion. It is evident also that with the realisation that the kinetic theory is a dynamical theory in the wider sense, the old conception of "solid" bodies has completely disappeared, though the outward semblance of them is still retained for statistical purposes in the form of the virtual volumes of molecules.

---

[1] If it be assumed that molecules are rigid bodies of various sizes and shapes, the extension to liquids of the kinetic theory encounters the difficulties clearly pointed out by Jeans in his *Dynamical Theory of Gases*, Chapter VI.

[2] See pages 168-170 of his *Collected Papers*. If, however, he had developed this original conception he would have been forced to admit that, as a general conception, the principle of conservation of *vis viva* is untenable. Nevertheless, he clung to the possibility of a kinetic explanation of both molecular attraction and gravitation, as appears in some of his papers.

In accordance with the assumptions just postulated, we can restate the fundamental gas laws in the following form:—*The virtual free space or intermolecular volume in a given gross volume of gas varies inversely as the internal bombardment pressure, and directly as the absolute temperature and the molar concentration at any given bombardment pressure.* In this form the gas laws, as will be shown later, can be extended to liquids and solids, and hold good regardless of the fact that actual molecules attract one another and are compressible to an extent which becomes quite appreciable at very high pressures.

Where liquids are concerned it is customary, and for good practical reasons, to take as a standard of reference, not the absolute number of moles present, nor the volume occupied by one mole (molar volume), but the number of moles present in one litre. If we call this number N, the *internal* bombardment pressure P, and the volume occupied by the molecules of one mole b, and extend the same nomenclature to gases, the gas equation becomes $P(1 - Nb) = N\ .082\ T$, or when any units of pressure and volume are used, $P(1 - Nb) = NRT$. When P is constant $1 - Nb$ increases in proportion to T, and it is only when the magnitude of Nb is a negligible proportion of 1 that N diminishes in proportion to T. When Nb is nearly equal to unity, N will only diminish very slightly with T. Similarly, with T constant Nb increases as P increases, and *vice versa*. When $P =$ one atmosphere, $T = 273^\circ$, and $1 - Nb$ may be taken as 1, as in a very perfect gas at atmospheric pressure, the value of N is $\frac{1}{.082\ T} = .0447$. Similarly, in perfect gas compressed at 0° C. to the same molar concentration as water at the same temperature, N would be 55.5, and $P(1 - Nb)$ would equal $55.5 \times .082 \times 273.1 = 1262$.

As compared with the equation $P(v - b) = n\ .082\ T$ the new form of equation changes the standard of volume reference to one litre, instead of the number of moles present (or the volume occupied by one mole when the equation takes the form $P(V-b)=RT$). The units of pressure and temperature are still retained, however. The greater convenience introduced by this change becomes evident when liquids and solids are dealt with; and the change places the nomenclature for gases on the same footing as that introduced by Mohr in connection with the volumetric solutions used in chemical analysis.

The new form of notation will henceforth be, in general, adopted, and the following lettering used:—

N, N_a, N_b, etc. = moles per litre of pure substances.

N_1, N_2, N_3, etc. = total moles per litre of mixed substances.

n, n_a, n_b, etc. = moles per litre of any substance present in a mixture.

$1-Nb$, $1-N_a b_a$, $1-N_b b_b$, etc. = corresponding virtual free spaces per litre (or proportions of free space) in pure substances.

$1-N_1 b_1$, $1-N_2 b_2$, $1-N_3 b_3$, etc. = virtual free spaces per litre in mixed substances.

P = bombardment pressure in atmospheres within the substance or mixture of substances.

P_a, P_b, etc. = partial bombardment pressures of different substances in a mixture.

p will in general be used for external pressure, including external vapour pressure; but P_o will denote osmotic pressure, and P_d osmotic deficiency, while p_e and p_d will be reserved for elevation and depression of vapour pressure.

With this notation the equation for a pure gas is $P(1-Nb)=N\,.082\,T$, and for a mixture of gases $P(1-N_1 b)=N_1\,.082\,T$. When the units of pressure and volume are not specified, R must be substituted for .082 in these equations, and N becomes the number of moles in unit volume. $1 - Nb$ is, of course, only the proportion of virtual free space.

PV stands not only for pressure in atmospheres multiplied by the volume in litres of one mole, but also represents, when b can be considered negligible, as in a gas at atmospheric pressure, the work done in expanding one mole of gas against atmospheric pressure, with no other resistance such as molecular attraction, to the volume V. The units in which the work is expressed are litre-atmospheres; but these can easily be expressed in any other units. With the new notation, the value of P represents the work in litre-atmospheres when Nb is negligible, and $P(1 - Nb)$ or $(1 - Nb)P$, when Nb is appreciable. The work is less when Nb is appreciable, because the expansion starts from the volume Nb and not zero volume. Similarly, n .082 T, or with the new notation N .082 T, represents the work in litre-atmospheres for n or N moles at T.

As standard atmospheric pressure = 1033.3 grammes per square centimetre, and the virtual free space in one mole of a very perfect gas is 22.39[1] litres at atmospheric pressure and 0° C., the expansive work represented by one mole of perfect gas at this pressure and temperature is 23,136,000 gramme-centimetres, or 231.36 kilogrammetres; and this equals .082 T (or 22.39) litre-atmospheres. Since the mechanical equivalent of one gramme-calorie is .4272 kilogrammetres the energy represented by the 231.36 kilogrammetres is $231.36 \div .4272 = 541.6$ gramme-calories. But according to the kinetic theory the heat-energy of expansion varies as T. Hence, as van't Hoff pointed out, this heat-energy can be expressed most conveniently as a multiple of T: and since at 273.1° absolute the value of this multiple for one mole of gas is 541.6 gramme-calories, the value of the multiplier is $541.6 \div 273.1 = 1.983$ gramme-calories. In other words, the heat-energy required for expansion against atmospheric pressure per mole of gas and per degree of absolute temperature is 1.983 gramme-calories. Van't Hoff, however, took 2.0 as a sufficiently close approximation when very great accuracy is not needed.

---

[1] To allow for the virtual volume of the molecules, 22.39 is taken in place of 22.41.

When b becomes appreciable it is not PV, but $P(V-b)$, or, with the new notation, $P(1-Nb)$ that stands for the energy required to expand the gas. Through the failure of van't Hoff to realise this he was misled in his conception of osmotic pressure, though he made brilliant use of the fact that what he erroneously took to be the value in a liquid of PV stands for expansive heat-energy.

It will now be evident that, using the new notation, we can state the fundamental gas equation as the following chain of equations:—$P(1-Nb)$ litre-atmospheres, or $P(1-Nb) \times 231.36$ kilogrammetres, or $P(1-Nb) \times 1673$ foot-pounds $= N\ .082\ T$, or $NT \times 1.983$ gramme-calories. In the particular case where $N = 1$ the N drops out in this equation, so that it represents the energy used up in expanding one mole.

Since the sizes of different molecules vary greatly, it follows that at very high internal pressures, when the value of $1 - Nb$ becomes much less than 1, different gases, liquids, or solids at the same temperature and pressure may contain very different numbers of moles per litre, and correspondingly different proportions of virtual free space. In other words, the pressure produced in a given volume per mole of substance, or per molecule, will be different, though constant for each variety of gas.

Let us now suppose that without alteration of the high pressure, P, or the volume or temperature, a given volume, at the internal bombardment pressure P, of a large-moleculed gas is substituted for an equal volume of a small-moleculed gas; and let N_s be the number of moles per litre, at the pressure P, of the pure small-moleculed gas, N_l the number for the pure large-moleculed gas, N_1 the number for the mixture, and n_l the number for the large-moleculed gas present in the mixture, so that $N_1 - n_l$ is the number for the small-moleculed gas in the mixture. On the dynamical theory as stated above, the pressure or "partial pressure" exercised directly by the $N_1 - n_l$ moles of the small-moleculed gas will be, not $\frac{N_1 - n_l}{N_1} P$, but $\frac{N_1 - n_l}{N_s} P$. Similarly, the directly-exercised pressure of the n_l molecules of the large-moleculed gas will be, not $\frac{n_l}{N_1} P$, but $\frac{n_l}{N_l} P$. This follows from the fact that in accordance with the dynamical theory, the internal bombardment pressure of a substance at constant external pressure and temperature depends on the ratio of number of moles or molecules to available virtual free space, apart altogether from whether the molecules are large or small. The virtual dead space representing the volume of the molecules themselves has no influence on the pressure which they exert, though at very high molecular concentrations it has a great influence on the gross volume which they occupy, and on the number of molecules present in a given gross volume. On the dynamical theory the large molecules, even if they are so large as to be visible, are perfectly elastic, and therefore transmit without loss the impulses communicated to them by small molecules, though these transmitted impulses do not come directly from the small molecules, and are counted to the large molecules. In the mixture and the pure gas each of the n_l moles per litre has the same mean translational energy, and therefore "inhabits," or keeps clear, the same mean virtual *free* space, in addition to the virtual *dead* space due to the larger size of its molecules. The partial pressure of the large-moleculed gas in the mixture is $\frac{n_l}{N_l}$ P, and not $\frac{n_l}{N_1}$ P, since the latter value would imply that each of the large molecules had a greater total or gross space attached to it in the mixture than in the pure gas. This is a matter of fundamental importance in the right understanding of solutions, as will appear later. In a gas at ordinary atmospheric pressure

N, N_1, and N_a have so nearly the same values that it does not matter which of these expressions we use; but it matters very greatly at high molecular concentrations.

It might easily be supposed that though this conclusion applies with sensible correctness to gases at low pressures it is not correct at high pressures, since it leaves out of account the interspaces left between the molecules when they are packed as closely as possible together. If we had a heap of cannon-balls packed as closely as possible together, it would still be possible to put in between them a considerable amount of sporting small shot without increasing the gross volume of the heap. This would still be the case to some extent if we left, along with the small shot, sufficient space for them to move about in with far greater velocity than that of the cannon-balls, in accordance with the *kinetic* theory. Thus we might, as it were, add a certain number of molecular small shot to a gross volume of molecular cannon-balls without increasing the gross volume of the latter.

A little consideration shows that on the dynamical theory this argument is wholly fallacious. If we add more molecules, however small they be, to a given gross volume of gas, we *ipso facto* increase its pressure in exact proportion to the number added per unit of virtual *free* space present in the gas. On the dynamical theory, the volume of the small-moleculed gas at a given temperature and pressure is just the same whether the large-moleculed gas is present or not. The molecules are not rigid bodies like cannon-balls and small shot; and from this standpoint the difficulties which would otherwise exist as to the deduction of Avogadro's law disappear. These difficulties are alluded to by Rayleigh in his introduction to Waterston's paper (Waterston's *Collected Papers*, p. 210).

Instead of expressing the partial pressure of the large-moleculed gas as $\frac{n_l}{N_1}P$, we can evidently express it as $\left(1 - \frac{N_1 - n_l}{N_s}\right)P$, or $\frac{N_s - N_1 + n_l}{N_s}P$. We can see from this expression that when N_s and N_1 have practically the same value, as is the case in a gas at atmospheric pressure, the expression becomes simply $\frac{n_l}{N_s}P$. When, however, the gas is at very high internal pressure, as in the case under consideration, the value of N_1 may be far less than that of N_s. As a consequence, the partial pressure of the large-moleculed gas may be far greater than $\frac{n_l}{N_s}P$. If, moreover, the molecules of the admixed gas were smaller, instead of larger, than the original gas, its partial pressure would be smaller, instead of greater, than that given by the expression just remarked on.

CHAPTER II

HEAT-ENGINES AND THE DYNAMICAL THEORY OF HEAT

The dynamical theory of gases threw an entirely new light, not only on the physical meaning of heat and temperature, but also on the mode of action of heat when by means of its application to a gas in a heat-engine mechanical power is obtained. In the light of that theory the heat-engine is evidently a molecular mechanism. Nevertheless it has not been from the standpoint of the dynamical theory of the gaseous and other states of matter that scientific treatment of the action of heat in heat-engines has been approached.

Though Watt had shown how to measure in definite units the work done by the steam in a steam-engine, and how to obtain automatically a graphic record of this work, and though the successive improvements which he introduced were based on thoroughly sound scientific principles, attempts at a radical scientific treatment of the mode of action of heat-engines may be said to date from the publication in 1824 of Sadi Carnot's short book *Réflexions sur la puissance motrice du feu.* Sadi Carnot was a young engineer officer, son of the famous French mathematician and War Minister who organised the Republic's victorious forces, and uncle of a future great French President. He seems to have been neglected, if not ill-treated, by those who were at the time in power, but fortunately turned his attention to one of the scientific aspects of his profession before his premature death from cholera in 1832.

Carnot pointed out that motive power is obtained from a heat-engine through the alternate expansion and compression of a working substance which may be water and steam, or gas, or even a simple liquid or solid. By the application of heat the working substance is made to expand against pressure, thus doing work and usually also falling in temperature. The compression is at a lower temperature, and the working substance is reduced to or maintained at the lower temperature during compression by the action of the condenser, which receives the heat wasted in cooling the working substance or keeping it cool during compression. Owing to this lower temperature the motive power, if any, expended during compression is less than that gained during expansion, so that there is a net gain in motive power, which is utilised in any way desired. In the condensing steam-engine and other heat-engines which may be imagined, the working substance is not thrown away at the end of expansion, while in the non-condensing steam-engine and other actual engines it is thrown away into the atmosphere; but he points out that this makes no essential difference, since the atmosphere acts as a condenser and receives the wasted heat. The essential characteristic of a heat-engine is that power is gained from it solely in virtue of the application of extra heat to the working substance, which at the end of each expansion and compression or condensation returns to exactly the same state as before the heat was applied, so that the same process may be repeated indefinitely, often without the supply and removal of anything but heat. Carnot's demonstration of this is fundamental.

He interpreted the mode of action of a heat-engine in the light of the then accepted "caloric" theory of heat. According to this theory heat or caloric is a substance which is "imponderable" or without weight, though since it is a substance it is indestructible. It exists, however, in both the "latent" and "manifest" forms. In the latent form it is no longer sensible, but by various means may be again rendered manifest and sensible by a process of "liberation." Thus in the evaporation of a liquid or the melting of a solid heat becomes latent, and again manifest on reversing these processes, as Black's experiments showed. Friction also, or compression of a gas, is capable of rendering latent heat manifest, or "liberating" it, while expansion of a gas renders it latent. It is an entire mistake to suppose that Rumford's or Davy's experiments on the production of heat by friction or compression could not be interpreted on the caloric theory; and the same is true of Joule's later experiments on the "mechanical equivalent of heat," though Joule himself, like Rumford, regarded heat as molecular motion.

Carnot explained the action of a heat-engine as follows. When the working substance is heated and then allowed to expand, doing work, heat becomes latent owing to the expansion, and the substance thus cools itself to a lower temperature. In the stage of compression at a lower temperature the heat which had become latent in the expansion stage becomes again manifest, but is absorbed by the condenser. Thus no heat is lost in the working of the engine, though heat is transferred from the source of heat to the condenser. The heat so transferred is, however, now at a lower temperature. Thus the doing of net work by the engine depends on a letting down of heat from a higher to a lower temperature, and not on any disappearance of heat, just as the doing of work by a water-wheel depends on the letting down of water from a higher to a lower level, but implies no loss of water.

We now come to what is perhaps at present the best-known part of Carnot's reasoning. He pointed out that if we could reverse the action of a heat-engine by applying to it the same motive power as was gained each time in its forward action, and by so doing restore everything to its state before the forward action started, there could be no more perfect means of working the engine in obtaining work from heat, the temperature limits being the same. For if there were a more perfect means of working the engine, then when we let it work forwards by the better method, and afterwards worked it backwards in the ordinary way, we should have on the completion of this double process an excess of motive power, although everything was in exactly the same state as before any heat was applied. By coupling two similar engines together we could, alternatively, obtain a continuous supply of heat under the same conditions. Thus we should be able to obtain a continuous supply of motive power (or heat) without expenditure or alteration of anything. Any such occurrence is, however, entirely contrary to all human experience, and would, in fact, be grossly contradictory of what afterwards came to be known as the principle of conservation of energy. Carnot's reasoning here is, indeed, nothing but a particular way of applying this principle.

He then proceeded to describe what he believed to be a reversible heat-engine, with air as the working substance. In this engine, which consists of an ordinary cylinder and piston (Figure 1) driving the engine, the air, at the temperature of a cold body, B, is first heated by compression to the temperature of

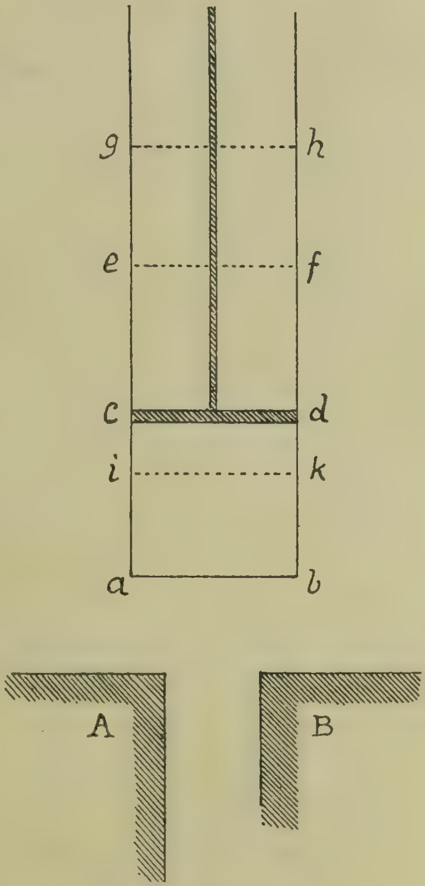

Fig. 1.—After Carnot.

a hot body, A, which serves as an external source of heat. The compressed and hot air is now allowed to expand isothermally in contact with this source of heat, while the temperature of the air remains that of the source during the expansion. The source of heat is then removed, and the air allowed to expand further, until, as a consequence, its temperature falls to that of the cold body. The cold body, B, at this temperature is next applied, and the air compressed isothermally, remaining at the same temperature, until it reaches its original volume. In the cycle there is a net gain of motive power, since the isothermal expansion is at a higher temperature than the isothermal compression. It is evident also that if an engine worked in this way, it could be worked backwards by applying the same net amount of motive power as was gained in working it forwards. It would draw as much heat from the condenser as had been deposited in the forward working, and deposit as much heat in the source as had been drawn from it in the forward working. This is the famous Carnot cycle; and upon Carnot's reasoning with regard to it an imposing structure of thermodynamical reasoning was afterwards, as we shall see, erected, and still forms a part of orthodox thermodynamical teaching.

Figure 1 is Carnot's original figure, and the lines *cd*, *ik*, *ef*, and *gh* show the successive positions of the piston. Figure 2 is a graphic representation of the pressure and volume of the air in the cylinder in accordance with Carnot's theory, the pressures and volumes corresponding as closely as possible to the

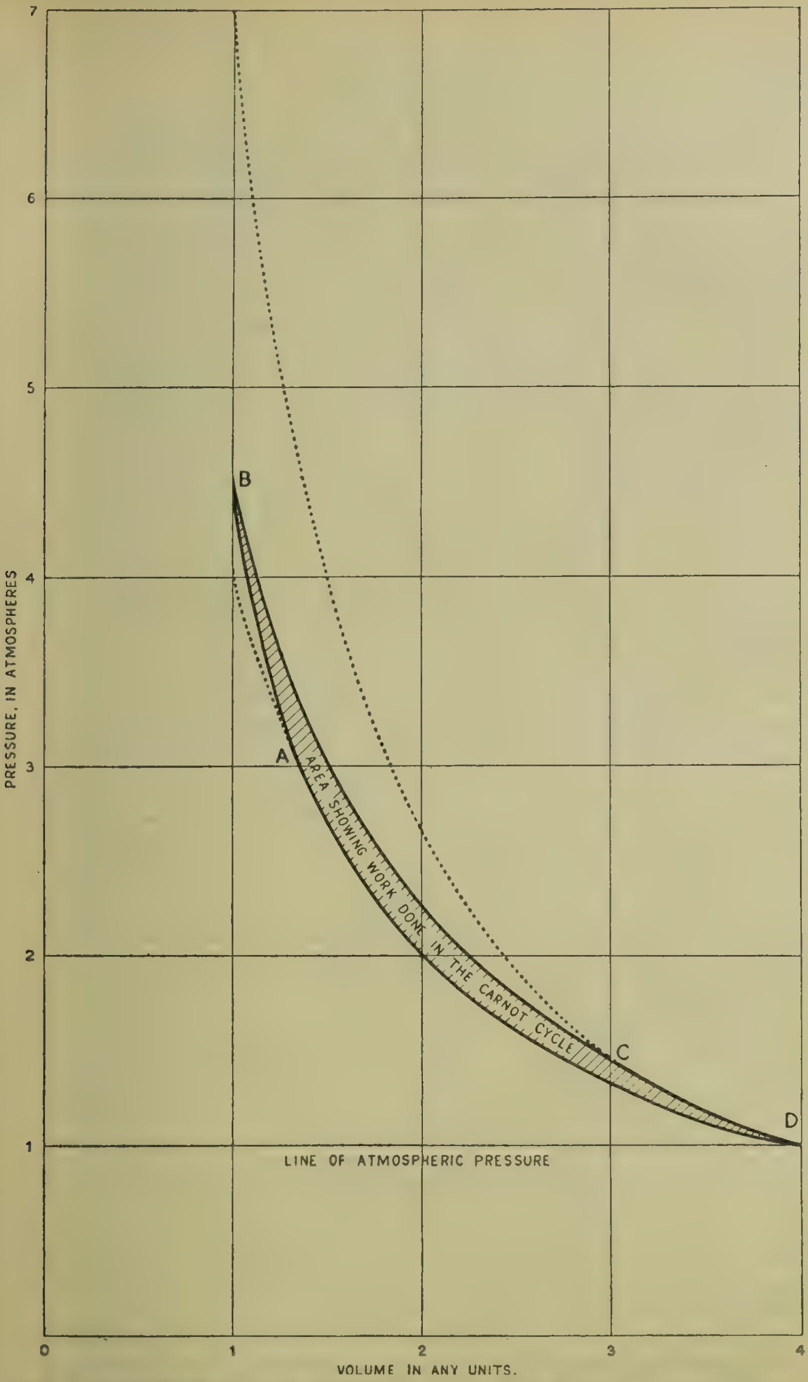

FIG. 2.—Showing the pressure and volume and the net work done, indicated by the area A B C D, in a Carnot cycle. The dotted lines indicate what would happen if in the forward working the isothermal compression were continued to the volume B, or in the backward working the adiabatic compression were continued to the same volume. In calculating the curves of adiabatic compression the value of γ has been taken as that for air.

positions represented in Carnot's figure, and thus giving a Watt indicator diagram A B C D of the net work supposed to be done per cycle by the engine.[1] It will be noticed that the cylinder is open at the top. Hence the pressure in the cylinder cannot fall below that of the atmosphere, and the pressure at A (corresponding to the position *cd* of the piston) must, by Boyle's law, have been three times the atmospheric pressure. From Carnot's description it follows that when the engine is originally started, before getting into its cycle, the pressure and temperature at A are atmospheric. In this case the temperature at A when the cycle is established must, by Charles's law, be thrice the absolute atmospheric temperature, or 546° C., if the atmospheric temperature was 0° C. The temperature of the "cold" body will thus also, according to Carnot's conception, be 546° C., and its absolute temperature will, as shown, be only one-ninth below that of the hot body, which will be 648° C. A line drawn vertically from A to the line of atmospheric pressure would show the start before the engine gets into its cycle.[2] It is tacitly assumed that no heat as such is either lost or gained by the working substance except that lost to the condenser or gained from the source of heat. Hence apart from this loss or gain the working of the engine would be what Rankine afterwards called "adiabatic." It is evident that out of the mechanical power developed during expansion only a part is available for work outside the engine. The rest must be stored up in a fly-wheel or other moving part, so as to furnish the power required for compression.

---

[1] Carnot did not know the law of adiabatic expansion, and thus drew his figure with impossible relations between the positions of the piston at different stages. Hence Figure 2 could not be drawn to correspond more than roughly with Figure 1.

[2] If we supposed that in starting the cycle the cylinder was filled with air at a pressure of three atmospheres, the cold body could be at atmospheric temperature, or 0° C., and the hot body at only 34° C.

When he wrote the paper sent to the Royal Society in 1845 Waterston had not seen Carnot's book, but he had seen a paper in which Clapeyron had endeavoured to represent graphically a Carnot cycle, and he expressed in a note (*Collected Papers*, p. 275) his approval of graphical representation. At the same time he pointed out (five years before Clausius did so) that the assumption that no heat disappears in the cycle of a heat-engine doing work is inconsistent with the dynamical conception of heat as developed in his paper. More heat must be taken up from the source than is thrown away into the condenser. He also re-drew Clapeyron's graphic representation of the cycle as Carnot had conceived it, since the dynamical theory of gases and heat had enabled him to draw curves of adiabatic expansion and compression, whereas Clapeyron did not know how to draw them, and had actually drawn them quite wrongly. Waterston had shown that adiabatic curves could be calculated on the dynamical theory from the ratio now usually designated γ, of the specific heat of air at constant pressure to its specific heat at constant volume. On similar reasoning it was shown later that the value of pv^{γ} is constant at every point on the curve. The contrast between correct graphic representations of Carnot's imaginary reversible cycle with various temperature limits and the crude representations of Clapeyron and of various current writers is a very remarkable one. The conventional lozenge-shaped indicator diagram of the imaginary cycle has little resemblance to the actual diagram.

In 1848 Kelvin, who was of course quite ignorant of Waterston's buried paper, or of how to draw the adiabatic curves, published an afterwards famous paper in the *Transactions of the Cambridge Philosophical Society.* In this paper he pointed out the fundamental importance of Carnot's reasoning. He was already familiar with Joule's experiments on the mechanical equivalent of heat, but quite logically did not consider that they were inconsistent with the caloric theory. He therefore continued to accept Carnot's reasoning and interpretations on the caloric theory; and he proceeded to develop mathematically the consequences of the reasoning as applied to the engine which Carnot believed to be reversible. He argued that since the doing of work by a heat-engine depends upon a letting down of temperature and not a disappearance of heat, and since the amount of motive power wasted during the stage of isothermal compression in the Carnot cycle depends upon how far the temperature of the condenser is below that of the source of heat, the efficiency with which a Carnot engine obtains motive power from a given amount of heat must reach 100 per cent. when the air cools itself by adiabatic expansion to absolute zero of temperature, and the condenser is at this temperature. In other words, he concluded that there is neither throwing away of heat nor of attainable motive power when the working substance of the Carnot engine has cooled itself to absolute zero and the condenser is at this temperature. The percentage efficiency of a Carnot engine thus gives us a thermodynamical scale of absolute temperature. If T is the absolute temperature at which heat is supplied to the engine during isothermal expansion, and T_1 the absolute temperature of the condenser, the efficiency of the engine is $\frac{T - T_1}{T}$, so that $T_1 = 0^\circ$ when the efficiency is 100 per cent.

The next step in the actual historical development of this reasoning was taken by Clausius in 1850,[1] and a few months later by Kelvin himself independently.[2] The experiments of Joule, demonstrating that a certain amount of motive power is needed to "liberate" a certain amount of heat, and the writings of Mayer, and later Helmholtz, on the conservation of energy, had begun to penetrate the mathematical screen behind which the naked unintelligibility of the caloric theory was concealed. Both Clausius and Kelvin adopted the view that heat is a form of kinetic energy, though they were still quite ignorant of the dynamical theory of gases. They adhered to Carnot's reasoning about his reversible engine; but the conclusion that heat is a form of kinetic energy carried with it, as they now pointed out, the conclusion that when a heat-engine does work, an amount of heat which is equivalent to the net work done in the cycle of the engine disappears, so that less heat passes into the condenser than is communicated to the working substance. Since, however, Carnot's reasoning about his reversible engine still appeared to them to hold good apart from his mistaken interpretation on the caloric theory, it seemed to follow that the efficiency of a Carnot engine is proportional to the letting-down of heat from the higher temperature in the source to the lower temperature in the condenser, and that Kelvin's previous conception of absolute temperature as being defined by the percentage efficiency of a Carnot engine still held good. It also appeared that heat acquires "thermodynamical potential" in proportion to the absolute temperature at which it exists, as compared with the absolute temperature of the condenser.

---

[1] Clausius, *Poggendorff's Annalen*, 1850.

[2] Kelvin, *Transactions of the Royal Society of Edinburgh*, xx., 261.

Waterston had interpreted absolute zero as the temperature at which the molecules of a gas are at rest relatively to one another. He seems to have objected strongly to Kelvin's interpretation of absolute zero, and to his thermodynamical reasoning generally, but was apparently unable to obtain the publication of a paper criticising Kelvin's deductions, which had come to be generally accepted. A manuscript which Waterston mentioned in his will, and which probably contained his criticisms, could not be found after his death from an accident. Let us, however, examine Carnot's and Kelvin's reasoning in the light of the dynamical theory of gases and temperature as formulated by Waterston.

The objections to Kelvin's thermodynamical scale of absolute temperature are of two kinds. In the first place such a thing as a reversible heat-engine is not known; and the demonstration will be given below that such an engine is not even a theoretical possibility.

In the second place, even if we assume that a reversible heat-engine is theoretically possible, the calculation of its efficiency with various drops in temperature assumes Boyle's and Charles's laws, of which no physical explanation is given. If we assume Charles's law, it gives us at once a definition of absolute temperature, and in a far more direct manner than when the matter is complicated by introducing the conception of a reversible engine. But the dynamical theory of gases gives us at once both Charles's law as applied to a perfect gas, and a clear physical explanation of it and of what absolute zero means. It also reveals to us the molecular mechanism by which a heat-engine works.

The accepted theory of the action of any kind of heat-engine has been based for so long on the conceptions of Carnot, Kelvin, and Clausius that probably few persons will at first be inclined to re-examine these conceptions. The matter is, however, one of very great importance, both practically and scientifically; and the fact that so great a physicist as Waterston was in disagreement with the orthodox theory points to the need of a re-examination.

We can put Carnot's argument, as finally summed up on page 38 of his book, into the form of a syllogism as follows: A heat-engine working between two temperatures cannot work more efficiently than when it works reversibly in Carnot's sense: the Carnot heat-engine is reversible: therefore there can be no more efficient engine working between two temperatures than a Carnot engine.

Unless both of the two premises are correct, it does not follow that a different kind of heat-engine, working between the same temperatures, cannot work more efficiently than a Carnot engine. It will be shown in the sequel that existing kinds of heat-engine can, as a matter of fact, work far more efficiently between two temperatures than a Carnot engine. The conclusion ought to have been "therefore it cannot work more efficiently between two temperatures." Meanwhile, however, we may confine the discussion to the Carnot engine itself.

For the reasons given by Carnot, and reinforced by the general formulation of the law of conservation of energy, the first premise cannot be attacked; but it would be preferable to put it into the form that if a heat-engine is working reversibly, it cannot work more efficiently between two temperatures than it does.

As regards the second premise, the position is very different. On the dynamical theory of heat and temperature, as formulated by Waterston, the doing of work by the Carnot engine during the expansion stage is due to the impact of innumerable perfectly elastic molecules on the perfectly elastic surface of the piston. Each molecule returns from its impact against the moving piston with less kinetic energy than it possessed before the impact. Hence the layer of working substance next to the piston is cooled, and the cooling extends to a constantly diminishing extent to the source of heat. During isothermal expansion the temperature of the gas cannot be even at the moment. It is only the mean temperature which is even, and which determines the pressure of the gas with any given mean concentration of its molecules. But in isothermal expansion this mean temperature must be lower than that of the source, otherwise no heat would pass, since where an innumerable number of molecules in chaotic movement at the same temperature are concerned, the chances are practically infinite that no measurable heat-energy would, on a balance, pass. Any heat that did pass would pass as readily from the working substance to the source as from the source to the working substance; and if the working substance were hotter than a body in contact with it, heat would necessarily pass into, and not from, that body. Conversely, during isothermal compression the temperature of the working substance, with heat constantly being communicated to it by the motion of the piston, must be above the temperature of the condenser. It follows, therefore, that the cycle of the Carnot engine is not really reversible in Carnot's sense, since if it were reversed more work would have to be done by the engine during isothermal compression than was done on it in the forward working during isothermal expansion; and during isothermal expansion less work would be done on the engine than by it during isothermal compression in the forward working. It also follows that the faster the piston is doing work, and thus expending or accumulating the heat in the gas, the greater must be the difference in temperature between the gas and the source of heat in expansion, and gas and condenser in compression, if expansion and compression are to be isothermal. The further, therefore, will the engine be from being reversible.

The theoretical irreversibility of the Carnot cycle, fully realised by Clerk Maxwell, is generally acknowledged. It is argued, however, that if the engine were working slowly enough the necessary difference in temperature between source or condenser and working substance could be made so small that the engine would be, practically speaking, reversible. This was not Carnot's idea; but let us examine it with the whole molecular mechanism of a Carnot engine in view, and also bearing in mind the conditions under which the engine works.

We can make the isothermal expansion very slow if we please by letting the piston work against a great enough inertial resistance. But if this resistance is outside the engine, and not due mainly to the existence of a heavy fly-wheel or its equivalent as part of the engine itself, there will, at the end of expansion, be insufficient energy available for the work of compression, so that the Carnot engine will stick, and its cycle cannot be completed. If, on the other hand, no external work were done, though, in consequence of adiabatic expansion and the temperature of the condenser being low, the isothermal compression (now possible, but very slow owing to the heavy fly-wheel) was at a lower temperature than that of isothermal expansion, the engine, as Carnot imagined it, would have a gain of velocity at the end of each completed double operation, and would thus not be working in a cycle in Carnot's sense. Its velocity would go on increasing until (if we neglect the friction of the engine and loss of heat by it to its environment) the difference in temperature between the isothermally expanding gas and source, and that between the gas which was being compressed isothermally and the condenser, became so great that the temperature of the gas during compression was the same as that during expansion. Only then would the theoretical engine begin to work in a cycle, and would its velocity, in consequence, cease to increase, whether the fly-wheel was heavy or light. All the heat communicated to the gas during expansion would then be passing into the condenser during compression. It is also evident that the greater the difference in temperature between source and condenser, the faster would the engine have to be moving before it ceased to accelerate. With a small difference in temperature, on the other hand, a correspondingly slight increase in velocity would stop the acceleration, so that the engine could only move very slowly, even though it was doing no external work.

With a very heavy fly-wheel the first complete double stroke of the engine is very slow, so that it might seem that the temperatures of the gas during expansion and compression would be very close to those of the source of heat and condenser respectively. If, now, we let the engine do just so much external work during this double stroke that no net acceleration remained at the end of compression, we might seem to have an engine working in a cycle, and yet very nearly reversibly. Its thermal efficiency would seem to vary very nearly as $\frac{T - T_1}{T}$, when T is the temperature of the source, and T_1 that of the condenser.

But in this reasoning we have overlooked the fact that owing to the cylinder walls being composed, just like the gas, of molecules and atoms, heat is inevitably leaking from and to the gas into and out of the walls, as well as from them to the outside so long as their temperature is above that of the environment. The source of heat itself is also inevitably losing heat. Even if, by very thorough lagging, we greatly diminished the rate of leakage to the outside, the cylinder walls and lagging would take up heat from the gas while its temperature was raised, and give off heat again to the gas while its temperature was lowered. As a result of this, adiabatic or isothermal expansion or compression is impossible with a slow-running engine; there would only be a very small mean difference in temperature between expansion and compression, so that very little net work could be done in the cycle; and very large amounts of heat would pass into the engine in each cycle and leak uselessly out of it into the condenser and to the outside. With adiabatic conditions as regards the cylinder walls, source of heat, and the bottom of the cylinder except during isothermal expansion and compression, we could, if the engine was moving slowly enough, calculate the efficiency approximately from the difference in absolute temperature between source and condenser, since the efficiency would be given approximately by the fraction $\frac{T - T_1}{T}$, when T and T_1 are defined as by Kelvin. In actual fact, however, no such efficiency could by any possibility be realised. The engine, however well the cylinder was lagged, would probably stick owing to friction before completion of its prolonged and extremely ineffective and wasteful cycle.

The loss of power owing to alternate heating and cooling of the cylinder and piston was clearly recognised by Watt, who tried in vain to remedy it by making a cylinder of wood soaked in oil, and minimised loss of heat to the outside by jacketing the cylinder. The point which it is necessary to make here is that the combined loss, though not either separately, is unavoidable, and tells enormously when the engine is moving very slowly, so that there is no possibility of the engine being even approximately reversible. If, on the other hand, the piston is moving fast the engine will be quite irreversible owing to the cause already pointed out and generally acknowledged. If he had lived to read Carnot's book, Watt would probably have questioned at once the possibility of a reversible engine.

It thus appears that whether the engine is working slowly or quickly, and however effectively it, and the source of heat, may be jacketed, its action in converting heat into mechanical energy is completely irreversible in Carnot's sense, and its thermal efficiency is correspondingly limited, quite apart from the limit due to restriction in the relative difference in absolute temperature between source and condenser. Under no actual circumstances can the efficiency of the engine be $\frac{T - T_1}{T}$ when T and T_1 are given the meaning assigned to them by Kelvin. Nor can the efficiency of a Carnot engine, varying, as it does, according to the rate at which the engine works, give us a scale of absolute temperature. Reasoning which, like the reasoning of Carnot as regards reversible engines and its mathematical elaboration by Kelvin and his followers, does not take into account conditions which necessarily exist, has no application to the actual world or actual heat-engines. In so far as quantitative thermodynamical reasoning neglects conditions which necessarily exist, it is only moving in a world of meaningless mathematical symbols and graphical representations of them.

When we ask the ultimate reason why the Carnot engine is not reversible and why the rate of doing work is so all-important in a heat-engine, we can see very easily that the explanation lies in the fact that heat is not merely molecular kinetic energy, but chaotic energy, as Waterston pointed out. Let us suppose that, instead of being a chaotic swarm of very rapidly moving molecules, the working substance in the Carnot engine consisted of molecules flying with equal rapidity, but up and down between the bottom of the cylinder and the piston, without ever colliding with one another or with the walls of the cylinder except with the bottom during isothermal expansion and compression. Let us also suppose that in the penetration of heat through the bottom of the cylinder there was a similar absence of collisions between the molecules, except such collisions as gave straight up and down movements. It is clear that each molecule striking the rising piston would return with only very slightly less kinetic energy, and would have its energy restored each time it came into contact with the heated bottom of the cylinder. No heat would escape into or out of the cylinder walls. If we defined the absolute temperature of this peculiar working substance as being proportional to the kinetic energy of its molecules, its temperature would during isothermal expansion be uniform throughout, and even when the piston was travelling very rapidly would be only very slightly lower than that of the source of heat at the bottom, since the piston velocity would be only a small fraction of that of the gas-molecules. During isothermal recompression at a lower temperature its temperature would also be only very slightly higher than that of the condenser, so that a Carnot cycle with this working substance would be *practically* a reversible one, even though its rate of doing work was high. Such a gas would, however, imply the same kind of "co-ordination" as we seem to find in biological and conscious activity.

But now let us recollect that the working substance is an actual gas, with its chaotic conditions, demonstrated by the facts on which the dynamical theory of a gas is based: also that the cylinder walls must, so long as the gas is hotter than they are, take up chaotic heat-energy and pass it on: also that the heat-energy passing through the bottom of the cylinder and the walls of the source of heat is chaotic. The molecules striking the rising piston and cylinder walls will each lose a little energy, and will return, not into a layer possessing the same energy as those which have just been in contact with the source of heat, but into a layer which has already been denuded of energy through successive previous contacts of the molecules with molecules coming from the piston surface and cylinder walls with less energy. As we go further and further down in the layers of the gas it is only by slow degrees that the molecular energy will approximate to that of the bottom layer. In other words, the gas will be cooled throughout, as compared with the centre of the source of heat, the maximum cooling being in the layers next the piston and cylinder walls, and the minimum in the layer next the source of heat. The molecular concentrations in these layers will vary inversely as their absolute temperatures, the pressure being even throughout. In proportion as the gas near the piston and cylinder wall cools work will be done on it, and heat instantaneously communicated to it, by the rest of the gas in the cylinder. But this cause only diminishes the cooling which would otherwise exist. The work done on the piston will, with the same mean molecular concentration of the gas, depend on the mean temperature of the gas, which determines the pressure, and will thus be less for a given expansion than if the molecules possessed practically the temperature of the source of heat. Exactly similarly, the loss of heat to the cylinder walls will depend on how far the *mean* temperature of the gas is above that of the walls, since though the gas next to the walls is cooled, its molecular concentration is correspondingly increased. Thus the loss of heat to the walls is inseparably linked up with the consumption of heat in doing work.

If the piston is travelling at less than the optimum velocity, there will be excessive loss of heat and net work owing to leakage of heat into and out of the cylinder walls and the walls of the source of heat. If the piston is travelling too fast, the work done on the engine during expansion will be diminished owing to the lowered mean temperature of the gas, while that done by the engine during compression will be increased owing to the raised mean temperature of the gas, thus also causing excess loss of heat and net work.

In the non-chaotic "gas" first imagined the conduction would be instantaneous, and there would be no loss of heat to the cylinder walls; but in the actual gas it will be very slow, and accompanied by much loss to the cylinder walls, owing to the chaotic conditions. Thus there will always be a great mean difference in relative temperature between the top and the bottom and sides of the cylinder, whether the piston is travelling slowly or quickly.

These are consequences of the chaotic molecular conditions assumed on Newtonian principles to exist in any molecular substance. The kinetic energy communicated to the working substance from the source of heat is bandied about, up and again down, hither and again thither, into and again out of molecular spin, before it finally reaches the piston; and meanwhile the energy is being lost through the cylinder walls and those of the source of heat, which take up and transmit heat in a similar chaotic manner.

On carefully reading Carnot's book we can see what misled him. In the compression stroke of a condensing steam-engine the temperature of the condenser surface is or may be practically the same as that of the steam in contact with it, and in the isothermal part of the expansion stroke the water in the boiler is, or may be, at about the same temperature as the steam. In this case, however, heat is "liberated" or rendered "latent" at the condenser or boiler surfaces, and not at the piston surface. Carnot made the incorrect assumption that the liberation or disappearance of heat occurs in the same way in a gas-engine as in a steam-engine. He also, since he did not understand what heat is, failed to see that adiabatic working conditions in the walls of a heat-engine are impossible; and his successors, with the probable exception of Waterston, failed to realise the full bearing of the dynamical theory of the states of matter and of temperature on the mode of action of heat-engines.

We have only to picture the molecular chaos as Waterston described it, in order to see that in the light of the dynamical theory of the states of matter and of temperature there was a fatal flaw in Carnot's reasoning, and that the mathematical elaboration could not ultimately conceal the flaw. The mathematical elaboration applied only to an ideal world which has no relation to the actual world.

It will now be evident that a reversible gas-engine, such as Carnot described, with the source of heat and condenser at the same temperatures respectively as the working substance, is an impossibility in the light of the dynamical theory of gases and heat, and represents conditions which cannot be even approximately realised in actual engines. We can, however, imagine a gas-engine working like the Carnot engine in so far as the working substance is heated by compression in the absence of the condenser, and cooled by expansion in the absence of the source of heat, but allowed during the first part of its expansion to absorb heat from a source, while during a corresponding part of the compression heat is absorbed by the condenser. Though such an engine would, for reasons to be pointed out later, be unsatisfactory in practice, it might be capable of doing work; and we can proceed to discuss the theoretical thermal efficiency which it could reach with various amounts and rates of work done in the cycle, and with the source of heat and condenser at various temperatures. For convenience, we may call such an engine a Carnot engine, though it is not reversible in Carnot's sense.

In the light of the dynamical conception of heat, the thermal efficiency of a heat-engine is the ratio of the mechanical energy obtained from the engine to the heat-energy supplied, both quantities being regarded simply as energy, so that they can be stated in terms of either units of mechanical work or units of heat. If H be the heat applied per cycle or unit of time, h be the heat equivalent of the work obtained, and H_1 the heat thrown away, it is evident that the thermal efficiency is $\frac{h}{H}$ or $\frac{H - H_1}{H}$, since $h = H - H_1$.

From the foregoing discussion it is evident that if the Carnot engine were moving very slowly a great deal of heat would have time to leak into and through the cylinder walls and those of the source during one part of the cycle, and back out of them, as well as through them, during the rest of the cycle. This would bring the mean temperatures of expansion and compression very close together, and cause great waste of heat. The thermal efficiency of the engine would thus be hopelessly low, if, indeed, since friction would be unavoidable, it could ever complete a cycle. On the other hand, it is equally evident that if the engine were travelling fast enough the temperature of the gas during isothermal expansion would be so much below that of the source of heat, and the temperature of the gas during isothermal compression would be so much above the temperature of the condenser that the thermal efficiency would also be hopelessly low. Though the loss of heat into and through the cylinder walls and those of the source per cycle might be for the most part prevented by the faster travel, the direct loss to the condenser in each cycle would be greatly increased. If the engine were, by externally applied force, made to run very fast, it would, in fact, act as a very efficient brake, as in the case of a motor-car engine on low gear when the car is running downhill. A further point is that, other things being equal, the further the mean temperature rises in the cycle the greater will be the leakage of heat into and through the cylinder walls.

It follows that in order to obtain the highest thermal efficiency from the engine the frequency of its cycle must be neither too low nor too high for the difference in temperature between source and condenser. The optimum frequency will also depend on the size of the cylinder, since, with a similar shape of cylinder, the surface, at which heat communication occurs, increases only as the square of the diameter, while the capacity increases as the cube. With a similar size and shape of cylinder the optimum rate will also depend on the concentration of the enclosed gas. How much we can increase the cycle rate advantageously will likewise depend, quite obviously, on the difference in temperature between compression and expansion. If this difference is small, a quite small increase in piston rate would bring the temperature of compression up to that of expansion, and thus reduce the efficiency to zero. Hence the engine could only do work at a very slow rate, so that its horse-power would be very low. If, however, the difference is great, the rate could be greatly increased without bringing the relative temperatures of expansion and compression too close. The horse-power of the engine, apart from mere increase in thermal efficiency, could be increased greatly by increasing the difference in temperature between expansion and compression, and this is a matter of the greatest practical importance in a heat-engine.

In considering the actual thermal efficiency which a Carnot engine might attain it will be assumed provisionally that the working substance is a "perfect" gas, consisting of what Waterston called "free and perfectly elastic" molecules, not even subject to the action of gravitation: also that the engine itself is practically frictionless, and has a perfect vacuum above the piston. Under these conditions heated gas might, if the piston were moving fast enough to make the expansion practically adiabatic, continue to expand indefinitely, at the same time reducing its temperature by doing work at a rate which diminished till it became indefinitely small as the temperature approached absolute zero. In the cycle of a heat-engine, however, the piston rate is limited, and it would only, by wasting time, reduce the horse-power and would increase the friction, if the temperature during expansion were allowed to fall below that of compression in presence of the condenser.

It will be convenient to consider first how the efficiency would vary if there was no leakage of heat into and out of or through the cylinder walls, and no other accidental loss of heat, so that the conditions were perfectly adiabatic in the absence of communication of heat from the source of heat or to the condenser. The temperature of the gas in the cycle of the engine would be raised by adiabatic compression, and lowered to an equal extent by adiabatic expansion. Hence, in accordance with the dynamical conception of heat, the work done on the gas in adiabatic compression would be equal to that done by it in adiabatic expansion, so that no net work would be done in the adiabatic parts of the cycle. The net work actually done would be due simply to the temperature of the gas being lower during isothermal compression than during isothermal expansion. When, moreover, we spoke of isothermal expansion or compression this would mean that the average temperature had a certain value; for the rate of expansion or compression in a reciprocating engine is varying all the time, and this would affect the actual temperature of the gas at any moment.

Figure 3 is an approximate graphic representation of the pressure and volume of the gas in the cycle of the hypothetical engine, with the absolute temperature of the gas during isothermal expansion assumed to be twice that during isothermal compression. In drawing the adiabatic curves the value of γ (the ratio of the specific heat of the gas at constant pressure to that at constant volume) has been taken as the value for air; and the value of pv^γ is constant at all points on these curves, while the value of pv is of course constant at all points on the isothermal curves. It will be seen that the isothermal expansion is to double the volume, while the corresponding isothermal compression is to half. The area A D G H, representing

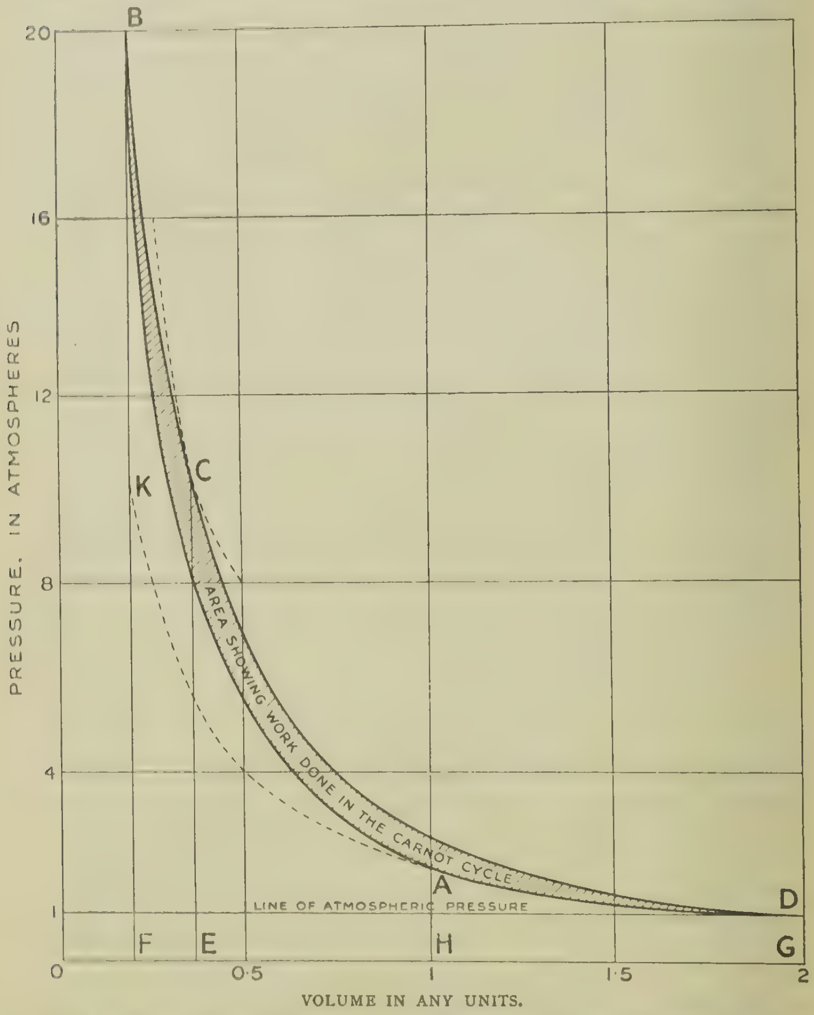

FIG. 3.—Carnot Cycle with Lower Absolute Temperature 50 per cent. below Upper.

the work done in isothermal compression, is half that done in isothermal expansion, represented by B C E F, since the absolute temperature in isothermal compression is half that in isothermal expansion, and the extent of the isothermal compression is, owing to the nature of the adiabatic expansion and compression, increased in exact proportion to the diminished molecular concentration of the gas. The area A B F H, representing the work done in adiabatic compression, must, of course, be equal to the area D C E G, representing the work done in adiabatic expansion,

There is thus, in the theoretical cycle, net available energy equal to half the heat-energy expended in isothermal expansion. In other words, the theoretical thermal efficiency of the engine is 50 per cent., and if we take T and T_1 as the maximum and minimum temperature of the *working substance* (not those of the source of heat and condenser), the thermal efficiency of the theoretical engine varies as $\frac{T - T_1}{T}$. A greater thermal efficiency would evidently be obtained if the temperature of isothermal expansion were higher, or that of isothermal compression lower. If, moreover, the movement of the piston were extremely slow the temperature of isothermal expansion and compression would become almost equal to those of the source and condenser, so that in this particular case we might take T and T_1 as respectively the temperatures of source and condenser. Thus with the absolute temperature of the condenser at 273°, or 0° C., we could, with a very slow rate of movement, obtain an efficiency of 50 per cent. with the source no hotter than about 273° C. With increasing speeds and the same temperatures of source and condenser the thermal efficiency would diminish, but the horse-power of the engine would at first greatly increase with increasing speed, and would continue to do so until the thermal efficiency was reduced to half in consequence of the increased speed. With still further increase of speed the horse-power, as well as the thermal efficiency, would gradually drop to zero.

As we have already seen, these calculations, and Fig. 3, which embodies them, represent a merely imaginary state of matters, and not anything which can possibly be realised in an actual heat-engine. What we have now to consider is the thermal efficiency and horse-power actually obtainable from a Carnot engine, with a given difference in absolute temperature between source and condenser.

With adiabatic conditions as regards the cylinder walls, etc., and an extremely slow rate of movement of the piston, the thermal efficiency of the engine would be $\frac{T - T_1}{T}$, where T was the absolute temperature of the source, and T_1 that of the condenser. We can make the engine work as slowly as we please, but we cannot possibly realise adiabatic conditions as regards the walls of the cylinder and source of heat; and the slower the engine works the more fatal is the latter fact to thermal efficiency. On the other hand, the faster the engine works the more is the thermal efficiency reduced owing to the approximation of the temperature of compression to that of expansion. It is clear also that with a given difference in temperature between source and condenser the maximum efficiency will be reached when the combined adverse influences of increase of rate and absence of adiabatic conditions are at a minimum.

When the engine is working so fast that the temperature of compression is equal to that of expansion the thermal efficiency is reduced to zero; and with any faster rate the engine would be acting as a brake, converting mechanical energy into heat instead of converting heat into mechanical energy. But when the point of zero efficiency owing to increased rate is reached, the adverse influence on the efficiency of imperfect adiabatic conditions is necessarily also zero, since the leakage of heat affects the work of expansion and compression equally. With a faster rate the leakage would diminish the work done in compression and increase that done in expansion, thus diminishing the brake action of the engine, and acting in an opposite direction to what leakage does at the slower rates.

There is thus an inseparable connection between the influence of rate of working and that of imperfect adiabatic conditions on the thermal efficiency which the engine would have if it could work very slowly and at the same time with perfect adiabatic conditions. When the former influence diminishes the thermal efficiency 100 per cent. the latter influence is necessarily zero; and as the rate diminishes the effect on thermal efficiency of the former influence diminishes from 100 per cent. to zero, while the effect of the latter influence increases from zero to 100 per cent.

Now when, other things being equal, we increase the piston velocity during isothermal compression, we increase the rate at which heat is communicated to the gas through the work done on it. We at the same time diminish proportionally the time available for the heat to escape into the condenser. But simultaneously the time available for heat to escape from the walls into the gas is correspondingly diminished. Thus the rise of temperature is due solely to the increased rate at which kinetic energy is being communicated to the gas. This rate increases as the square of the piston velocity. Hence the temperature of the gas increases as the square of the piston velocity.

Similarly, increase of velocity reduces the temperature of isothermal expansion in proportion to the square of the velocity. This inference presumes, of course, that the engine is working in a cycle. When this is the case, a decrease in external load, with consequent increase in velocity, implies a corresponding increase in the rate at which mechanical energy is communicated to the flywheel or its equivalent, and transmitted to the gas during compression.

At a certain critical velocity, V, the mean temperature during compression would, if the engine were frictionless, become equal to that during expansion. At this velocity the external work done would be zero, as would also the internal leakage through the walls, since there is no difference in mean temperature between them and the gas. The heat drawn from the source per cycle during isothermal expansion is diminished owing to the diminished time for communication of heat to the gas, along with the disappearance of internal leakage. But the heat absorbed per unit of time is increased in proportion to the greater difference in temperature between the gas and the source of heat. Thus the engine is absorbing more heat per unit of time than if it were working slowly. All this heat is, if direct leakage to the outside is inappreciable, converted into mechanical energy during expansion, and reconverted during compression into heat, which goes into the condenser and is thus wasted without any net work being done. As it is only with piston-velocities below V that any positive thermal efficiency is possible, it will be convenient to express velocity, v, as a fraction of V, treating V as unity.

In proportion as the temperature of the gas is higher during isothermal compression, and lower in isothermal expansion, the stages of expansion or compression in the absence of the source of heat or condenser must be shortened: otherwise neither expansion nor compression could, on an average, be isothermal. At the velocity V these stages disappear completely, so that the indicator diagram is equivalent to nothing but a single shortened line, representing isothermal, or virtually isothermal, expansion and compression at the same temperature.

Let us now consider on what the value of V, the limiting velocity, depends: for this is absolutely fundamental in the understanding of a heat-engine. In the running of the engine while it is doing net work, the walls, including piston, will take on a certain mean temperature, midway, if leakage directly to the outside is practically negligible between those of the gas in isothermal expansion and compression. During half the cycle, therefore, the walls will be cooling the gas, and during the other half warming it, although, owing to the heat capacity of the walls being relatively very large, their temperature may hardly vary. When the engine is running very slowly, so that the differences in temperature between the gas and the walls are at their maximum, the cooling effect of the walls during isothermal expansion, and warming effect during isothermal compression, will be at the maximum.

The stages of expansion and compression in the absence of condenser or source cannot possibly be regarded as actually adiabatic or isentropic. But if leakage to the outside has been made so small that it is not appreciable, the gas will lose heat to the walls during the first part of the expansion and last part of compression, and gain as much during the last part of expansion and first part of compression, so that for practical purposes we could treat the expansion and compression as adiabatic.

The mean temperatures of the gas during isothermal expansion and compression will depend upon the relative areas of the heating and cooling surfaces concerned, and their difference in temperature. In the Carnot engine the heating or cooling surfaces are alternately the bottom of the cylinder and the rest of its internal walls; and as the area of the bottom is small in relation to that of the rest of the walls, the difference in mean temperature between the gas in isothermal expansion and compression will, at its maximum, be much less than half the difference between T and T_1, if these are the respective temperatures of source and condenser. The adiabatic expansion and compression required are correspondingly shortened.

We can see at once from this consideration that the net work capable of being done per cycle by the Carnot engine is only a fraction of what Carnot and Kelvin imagined, and that the thermal efficiency of a heat-engine has no definite relation to the temperatures of either the source of heat or condenser, however slowly the engine may be working. It is singularly unfortunate that the temperatures of source of heat and condenser have come to play such a direct part in the theory of heat-engines. We can also see at once that the value of V for the Carnot engine will be low, since a quite small velocity would bring the temperatures of expansion and compression together. The engine could thus, even if it were frictionless, do only a little work per cycle slowly, although the temperature of the source might be so high as almost to melt the bottom of the cylinder. If, moreover, in order to raise to the maximum the work done per cycle, we reduced the velocity to a small fraction of V, the engine would waste a quite disproportionate amount of heat, owing to the increased time afforded for internal leakage to the condenser and leakage to the outside.

We could increase or reduce the leakage, and correspondingly reduce or increase the value of V, by making the cylinder longer or shorter in relation to its capacity, and could reduce V by only applying the heat and cold to a portion of the bottom of the cylinder, or (to a less extent) by applying either the heat or the cold to only a portion of the bottom. We could somewhat increase the value of V by cutting down the period of isothermal heating and consequent expansion; but this would be at the expense of reducing the work done per cycle. The more we raise the temperature of isothermal expansion, the greater, owing to the prolonged adiabatic expansion, becomes the area of walls exposed, and consequent leakage of heat. V is evidently the runaway piston velocity of a frictionless engine doing no external work, and for each individual engine this velocity is a fundamental characteristic, so that we cannot treat thermal efficiencies in the abstract. In actual fact no engine is frictionless; but a frictionless engine is at any rate a convenient provisional conception, corresponding to what may be called theoretical efficiency.

Let us now consider how the thermal efficiency of a frictionless Carnot engine would be influenced by altering the velocity, which can be effected by altering the load on the engine. The efficiency is $\frac{h}{H}$, when h is the value per cycle in heat-units of the mechanical energy obtained, and H the heat per cycle taken from the source.

If we made no allowance for leakage or variation in it with varying velocity, the thermal efficiency would be, if T_a were the temperature of the gas during isothermal expansion at so slow a rate that v^2 was very small, and T_b its temperature during isothermal compression at a similar rate, $\frac{T_a - T_b}{T_a}(1 - v^2)$, v being expressed as a fraction of V. This would, however, be equivalent to measuring $\frac{h}{H}$ with a leaky measure, without allowance for either the leakage or variation in it. $\frac{T_a - T_b}{T_a}(1 - v^2)$ represents not the thermal efficiency, but only the net proportion of mechanical energy finally obtained from that proportion of the heat taken in which is converted into mechanical energy during expansion. The actual amount of heat taken in is not only the amount converted into mechanical energy, but in addition the amount which leaks into the walls. Thus H consists of H_a, the heat converted into mechanical energy during expansion, and H_b, what leaks into the walls. Similarly H_1 consists of H_c, heat reconverted from mechanical energy during compression, and H_b, what leaks back from the walls into the gas (assuming the absence of appreciable loss to the outside). What leaks back during isothermal compression must be equal to what leaked into the walls during isothermal expansion, provided that the engine is working in a cycle and no heat is leaking to the outside. Thus the theoretical thermal efficiency is $\frac{(H_a + H_b) - (H_c + H_b)}{H_a + H_b}$, or $\frac{H_a - H_c}{H_a + H_b}$.

As the piston velocity alters, the values of H_a, H_b and H_c alter. With increasing velocity H_a and H_b diminish, while H_c increases; and the thermal efficiency depends on the ratio of $H_a - H_c$ to $H_a + H_b$. At the velocity 1 the efficiency is zero, since the temperature of compression is equal to that of expansion, so that H_c is equal to H_a and no net work is being done, though much heat is being wasted. The value of H_c has increased, and that of H_a has diminished, till H_c is equal to H_a; while, though the value of H_b has fallen to zero, that of H_a is still large, so that $H_a + H_b$ is still considerable. At a very low velocity, on the other hand, the value of $H_a - H_c$, which varies as $1 - v^2$, has practically reached a maximum at which H_a as well as H_c is constant, while the value of $H_a + H_b$ can still increase indefinitely with lower velocities, since, though H_a has practically reached a maximum, H_b can still increase indefinitely. The result is that with further diminution in velocity the efficiency comes nearer and nearer to zero, as it does when v becomes nearly 1.

$\frac{H_a - H_c}{H_a}$ represents the proportion converted into net work of the heat converted into work during isothermal expansion; and in an engine with a perfect gas as the working substance this proportion depends on difference in the absolute temperatures of isothermal expansion and compression, and equals $\frac{T - T_1}{T}$ when T and T_1 are the actual temperatures of isothermal expansion and compression. The ratio of $H_a + H_b$ to

H_a is a measure of the percentage by which the efficiency is diminished below $\frac{T - T_1}{T}$, and evidently this percentage will be high when v is either small or approaches 1, and will be lowest at an intermediate value of v.

Now the net work done per cycle diminishes as $1 - v^2$ as the piston velocity increases. On the other hand the internal leakage in proportion to this work diminishes as v increases, or as the time during which leakage occurs diminishes. Thus the net work done diminishes faster than in proportion to the diminution in leakage when the velocity 1 is approached, since $1 - v^2$ diminishes faster than v increases. The efficiency therefore diminishes. At a low velocity, however, $1 - v^2$ scarcely increases at all with diminishing velocity, while v progressively diminishes, so that here again the efficiency diminishes. At an intermediate velocity the efficiency is at a maximum, and at this velocity the proportion of heat lost by leakage must be equal to the proportion converted into net work, so that the thermal efficiency is 50 per cent. of $\frac{H_a - H_e}{H_a}$ or $\frac{T - T_1}{T}$. In other words H_b must be equal to H_a. But $\frac{H_a - H_e}{H_a}$ or $\frac{T - T_1}{T}$ varies as $1 - v^2$, while $\frac{H_a - H_e}{H_b}$ varies as $\frac{1 - v^2}{v}$. Hence equality can only be present when $1 - v^2 = v$, which is the case when $v = \cdot 618$ of V. At increasing velocities v increases more slowly than $1 - v^2$ diminishes, so that the efficiency diminishes, while at diminishing velocities v diminishes faster than $1 - v^2$ increases, so that the efficiency also diminishes. Thus at the velocity ·618 the efficiency is $\frac{T - T_1}{2T}$, and this is the maximum efficiency. The rate of work done with this efficiency depends, not on $\frac{T - T_1}{T}$, but on $T - T_1$.

With velocities increasing over 0.618 the net work done will diminish faster than the leakage, so that $H_a + H_b$ will increase in relation to H_a, and the efficiency will diminish. With velocities diminishing below 0.618 the leakage will increase faster than the net work done, so that $H_a + H_b$ will also increase in relation to H_a, and the efficiency will likewise diminish.

It should be clearly understood that since the internal leakage depends, not merely on the velocity, but also on the difference in temperature between the gas during expansion and during compression, which determines the net work done per cycle, the loss by this leakage is always a percentage of the net work done. This was assumed in the above reasoning. The loss by internal leakage varies, not merely as $\frac{1}{v}$, but as $\frac{1 - v^2}{v}$, and not only is the maximum value of $\frac{T - T_1}{T}$ only a fraction of what Carnot and Kelvin assumed for the Carnot engine, but the efficiency of the engine at its maximum, and without any allowance for loss from friction, is only half this fraction. We can also see that *from the indicator diagram we can* *draw no direct inferences at all as to the thermal efficiency.* The theoretical indicator diagram of a Carnot engine tells us the value of h, or $H - H_1$ but not that of H. When the value of $\frac{T - T_1}{T}$, as shown by the indicator diagram, is at its maximum the thermal efficiency is far lower than with a lower value of $\frac{T - T_1}{T}$, and is greatest when, owing to increased velocity, the value of $\frac{T - T_1}{T}$, and corresponding area of the indicator diagram, is reduced to less than two-thirds of its maximum.

The effect of leakage of heat direct to the outside has, so far, not been discussed. Such leakage, since the mean temperature of the cylinder walls over a cycle remains constant, varies with varying velocity as the time for it to occur, or as $\frac{1}{v}$. This is the same as internal leakage with a very low velocity, or as internal leakage when allowance is made for its reduction with lowered temperature difference. In so far, however, as external leakage exists, as it always must do however much it is reduced, it diminishes internal leakage, since by getting rid of heat during both isothermal expansion and compression it diminishes to a corresponding extent the loss of heat by internal leakage. Thus its existence makes no difference in calculating the efficiency with varying velocities, in so far as the existing value of $\frac{T - T_1}{T}$ enters into this calculation. But evidently external leakage, in so far as it is sensible, makes a difference in the existing value of $T - T_1$ at any velocity. As already pointed out, this value depends on the balance, in the particular engine, of the rate of supply of heat and the rate of leakage. If, owing to the external leakage, the temperature of the walls is lower than it otherwise would have been, the value of T is correspondingly lowered. The value of T_1 is also lowered, but to a less extent, since the temperature of the walls during compression is much nearer to that of the condenser than to that of the source of heat during expansion. Thus the value of $T - T_1$, and consequently that of V, the critical velocity is lowered. The value of $\frac{T - T_1}{T}$, or $1 - \frac{T_1}{T}$, is also lowered, as the reduction in T_1 is not in proportion to that in T. The engine therefore works slower and with less thermal efficiency, besides which a certain proportion of H_1, which would otherwise have been lost to the condenser, is now lost to the outside. This makes no difference to the relative efficiencies with varying velocities.

At an extremely low velocity the temperature of the walls would rise during expansion to a point considerably above their mean temperature, and fall during compression in a corresponding manner. Internal leakage would thus become considerably greater at the beginning than at the end of expansion in contact with the source, and passage of heat into the condenser considerably greater at the beginning than at the end of compression in contact with the condenser, so that neither expansion nor compression could be actually isothermal, even if the piston were moving with constant velocity. More and more of the heat which accumulated in the walls during expansion in contact with the source would have to be got rid of during compression, which would therefore start at a higher and higher temperature, while expansion would start at a lower and lower temperature. But evidently this tendency would be checked through leakage to the outside being less or greater in proportion as internal leakage was greater or less. It will be shown, however, that friction sets a lower limit to the velocity of an engine working in a cycle.

The external leakage might be so great that the mean temperature of the walls was below that of the gas during isothermal compression in presence of the condenser. It would still be the case, however, that owing to extra heat absorbed by the walls during isothermal expansion the temperature of isothermal compression would be higher.

If there were just sufficient external leakage to balance the heat which would otherwise pass into the condenser, the condenser itself could evidently be dispensed with without raising the mean temperature of compressive, or lowering the mean temperature of expansive work. In this case there would be no rise of temperature in the last stage of compression, and all pretence of a stage of isothermal expansion, or of adiabatic compression or expansion, would disappear, since the rise of temperature in the cycle would be produced solely by the application of the source of heat. The engine would, however, still do as much work as before, since the mean temperature of expansive work would be as much above that of compressive. The whole of the heat passing into the working substance from the source is either converted into mechanical energy or leaks to the outside; and as what leaks to the outside is equivalent to what would have passed into the condenser if there had been no leakage to the outside, the amount converted into mechanical energy is the same as what would have been converted with the help of the condenser. In other words the thermal efficiency is the same, and we have an engine in which the function of the condenser is distributed over the external walls, which are performing this function throughout the whole of the cycle. In the internal combustion engine and other possible engines we have similar conditions, as will be shown below; and condensers acting only during compression can evidently be dispensed with. Provided that the leakage to the outside, or supply of heat, is properly regulated so as to keep the maximum value of $\frac{T - T_1}{T}$ from falling, and the piston velocity is also properly regulated, we can secure maximum thermal efficiency without a special condenser acting only during one part of the cycle. How important these considerations are in connection with the steam-engine will be pointed out in Chapter VIII.

We are so accustomed to think of heat-engines in the light of Carnot's assumption that reversible heat-engines are theoretically possible, and not in the light of the dynamical theory of heat, that it is difficult at first to cut ourselves adrift from this inherited idea and its mathematical development. There seems, however, to be no alternative to abandoning the ideas and calculations based on Carnot's assumption.

Whether the efficiency of a heat-engine is connected, as in gas-engines, with the fraction $\frac{T - T_1}{T}$, or, as probably in muscles, is connected with temperature in a different manner, the conclusion that the maximum theoretical efficiency cannot exceed 50 per cent., and is dependent on the velocity of the cycle, seems absolutely fundamental in the theory of heat-engines, whether these engines depend on expansion by heat or are of other types.

The connection between loss of efficiency by leakage into and out of the cylinder walls and loss due to increasing piston velocity is evidently independent of the difference in temperature between source and condenser, or the absolute velocity required to extinguish the efficiency. In proportion, however, as the relative difference between T and T_1 is increased, not only does the maximal thermal efficiency increase nearer and nearer to 50 per cent., but the value of V also increases simultaneously, as well as the work done per cycle. Thus we gain in horse-power far more than we gain in thermal efficiency. With a small difference between T and T_1 the horse-power is insignificant. By lowering T_1 we cannot gain as much in horse-power as by raising T, since though the efficiency is increased there can only be a limited increase in the value of V and the work done per cycle.

No definite account has been taken as yet of direct loss of heat from the source to the outside, or of losses due to friction. In internal combustion engines direct losses of heat otherwise than through the walls of the engine are avoided, but losses from friction are common to all heat-engines.

Direct losses of heat from the source to the outside are, per cycle of the engine, evidently proportional inversely to the frequency of the cycle. With the source at a constant temperature, therefore, and the engine running at a velocity which gives maximum efficiency, the leakage from the source will have just the same influence as if a certain amount were taken off the value of $\frac{T - T_1}{T}$. If, now, we raise the value of $\frac{T - T_1}{T}$ by raising the temperature of the source, we correspondingly, other things being equal, raise the rate of leakage from the source; but as we also raise, not only the work done per cycle, but the rate at which the piston moves with maximal thermal efficiency, it follows that the leakage from the source tells less and less seriously as the value of $\frac{T - T_1}{T}$ is raised.

Loss due to friction is of a different character. It is a dead loss in the net available work done by the engine, proportional to the time taken per cycle, and increasing, not necessarily with velocity, but only with pressure on the bearings, provided that they are kept properly lubricated and cool. Thus it is a loss on h in the fraction $\frac{h}{\mathrm{H}}$, and this may be very serious if the value of $\frac{h}{\mathrm{H}}$ is low. Inasmuch, moreover, as the moving parts of an engine are heavy, there is a minimum pressure on all bearings, and therefore a minimum adverse pressure to be overcome, however small may be the work done, and consequently the average extra pressure on the bearings, during the cycle. In consequence of this an engine designed to work with high pressure differences cannot complete a cycle at all unless the relative difference in pressure between expansion and compression is considerable. In a reciprocating engine there is, in addition, heavy pressure of the piston rings on the cylinder walls during the whole of the cycle, and with increasing velocity rapidly increasing frictional loss due to the inertia of heavy reciprocating masses.

We can see at once from these considerations how heavily handicapped a reciprocating heat-engine with any working pressure is against a turbine, and in particular how heavily handicapped a Carnot engine, with its exaggerated length of stroke, would be (if anyone ever succeeded in making a Carnot engine which could complete a cycle without being pushed) against any ordinary form of reciprocating engine. Owing to the attenuated shape of its indicator diagram, the low piston velocity which it can reach, and the very high pressure existing at the point of maximum compression, the Carnot engine works with a maximum of friction; and the heat produced by this friction is just a negative value set off against the small proportion of heat capable otherwise of being converted into available mechanical energy. As, moreover, we increase the difference between T and T_1 in a Carnot engine, we must correspondingly increase the length of stroke, so that friction and leakage of heat into the walls increase in proportion to increase in work which could otherwise be done.

Since, particularly in the reciprocating engine, the heat produced by friction is, even when the engine is doing no external work, a dead deduction from the external work otherwise capable of being done, it follows at once that in so far as in any part of the cycle of an engine less net mechanical energy is being obtained than corresponds to the heat wasted in friction, that part of the cycle can with great advantage be cut off as being worse than useless, besides adding to the size and cost of the engine. This is one reason why, in actual heat-engines, the theoretical cycle of the engine is more or less curtailed, the corresponding amount of heat being thrown away into the condenser, and why in reciprocating engines the curtailment is specially heavy. We can also see why an engine designed for high pressure will be uneconomical, or fail entirely, with a low pressure.

As, other things being equal, we increase the mean difference between the pressure in the expansion and that in the compression stroke, the loss by friction increases, it is true; but only in proportion to the increased pressure, and not to the increased rate of work possible. Hence the proportional loss of energy per cycle becomes less at first, so that the possible net thermal efficiency increases more than in proportion to $\frac{T - T_1}{T}$. With less than a certain value of $\frac{T - T_1}{T}$ an engine designed for higher pressure differences would not do any net work at all unless the stroke was shortened; and with a still lower value of $\frac{T - T_1}{T}$ it would simply stick in the compression stroke, however much the length of stroke was shortened.

An engine working at very much less than a certain piston velocity would begin to lose so excessively from friction that further diminution in velocity would evidently become impossible owing to sticking at the end of the compression stroke. Loss of energy by friction, just like loss of energy by ordinary leakage of heat, is a consequence of the fact that the walls of an engine are composed of molecules, to which kinetic energy is necessarily communicated in the working of the engine. A larger and larger proportion of this heat will escape by conduction the slower, other things being equal, the engine works. In the theory of heat-engines we can no more neglect friction than we can neglect ordinary escape of heat by leakage. A frictionless Carnot engine is thus in reality as much of a mere abstraction as a reversible Carnot engine.

The effect of increasing the velocity up to its optimum value is even more important as regards increase of horse-power than of efficiency; for not only can the engine do its work per cycle more efficiently, but it can also work a great deal faster. This is an immense practical advantage from various points of view. It seems clear that in the past far too exclusive attention has been paid in the academic theory of heat-engines to the mere value of the fraction $\frac{T - T_1}{T}$.

If a cylinder consisting of a substance which neither absorbed, nor conducted, nor radiated heat could be found, the loss of efficiency owing to heat leakage through the walls could be avoided, and we could slow down the rate till the efficiency corresponding to $\frac{T - T_1}{T}$ in Kelvin's sense was attained. But the existence of such a substance is inconsistent with the dynamical theory of heat as chaotic kinetic energy. We can, it is true, arrange cylinder walls which conduct heat badly; but if so the temperatures of both the surface of the walls and the gas will only fluctuate more during each cycle. Owing to its chaotic nature, we cannot stop the heat energy from escaping uselessly into, and again out of, the cylinder walls. We might imagine a Carnot engine working at temperatures not far from absolute zero, and with walls consisting of diamond or some other substance which took up no heat at these temperatures. Such a possibility points to ultimate limitations in the dynamical theory of heat.

In order to obtain satisfactory efficiency with heat escaping rapidly through the walls, as when a cylinder is cooled by a water-jacket, we require to increase both the piston velocity and rate of heat supply correspondingly. By increasing the speed and heat supply we not only obtain a greatly increased efficiency, but also a still more greatly increased horse-power, without increase in the size of engine. The importance of these considerations in connection with both the internal combustion-engine and steam-engine will appear below. With a Carnot engine, however, we should be powerless to increase the rate of heat supply sufficiently to meet increased leakage.

Even with the condenser and temperature of environment at absolute zero, the maximum theoretical efficiency of a Carnot engine would be far below 50 per cent., without allowance for direct loss from the source and for friction, which losses would be very heavy. Fig. 3 represents much more than the theoretical efficiency which could be obtained under this condition. It is evident, however, that if we define T and T_1 as the mean temperatures of the gaseous working substance in contact respectively with the source and condenser, the theoretical maximum efficiency of the Carnot engine will be $\frac{T - T_1}{2T}$; but the values of T and T_1 will be very different from those in Kelvin's sense.

If the engine could work adiabatically except when the cylinder was in contact with the source or condenser we could, without loss of efficiency, reduce the velocity indefinitely, and it would be necessary, in order to obtain maximum efficiency, to increase the expansion stroke as the temperature of the source rose above that of the condenser. In an actual engine, however, it would be worse than useless to continue the expansion so far that the temperature of the gas fell below the initial temperature of compression in presence of the condenser. As the latter temperature would be considerably above that of the condenser, and the more so the higher the temperature of the source, the expansion can be kept low accordingly; and if it were not thus kept low much time would be wasted unnecessarily in the cycle, so that the horse-power and efficiency of the engine would be greatly reduced owing to waste of time and increased friction.

The indicator diagram of an ideal Carnot engine working with maximum efficiency would be different from Fig. 2 or Fig. 3, since no part of either the expansion or compression would be either quite adiabatic or quite isothermal; but as the leakage of heat during what would otherwise be adiabatic expansion or compression would be in opposite directions during the first and second halves of what would otherwise be adiabatic cooling or heating, there would not be much change in the diagram for these phases; and owing to the nearly steady temperature of the cylinder walls the rest of the expansion or compression would be nearly isothermal, in so far as isothermal expansion or compression are possible in a reciprocating engine.

Let us now suppose that the piston of the Carnot engine was exposed above to the atmosphere, instead of to a vacuum. Owing to the work done on the atmosphere the work done otherwise by the engine would be reduced, but during compression the work done by the atmosphere would reduce the work to be done otherwise by the engine. In the light of the dynamical theory of gases we can, however, easily see that the work done on the air by the piston in expansion would be greater, though only a very little, than the work done on the piston by the atmosphere during compression. The temperature of the adjacent air would be slightly raised during expansion, and lowered during compression. In other words, there would be a very slight net loss to the engine owing to what we call the "friction" of the air.

In the above calculations as to the efficiencies of a Carnot engine with various differences in temperature between isothermal compression and expansion, it has been assumed tacitly that for the perfect gas assumed to be the working substance the value of γ is the same at different temperatures. In actual gases the specific heat at constant volume increases with temperature, so that the value of γ diminishes. As a consequence, more work has to be done on the gas during the last part of compression, but the gas does correspondingly more work in the last part of expansion, so that the maximum thermal efficiency, neglecting the effects of friction, is still $\frac{T - T_1}{2T}$.

The conditions on which the thermal efficiency of a Carnot engine depends have been discussed at considerable length. This is partly because the Carnot engine has played such a prominent part in the generally accepted theory of heat-engines, but partly also because the principles which have been applied in the discussion of the Carnot engine are applicable also to all other heat-engines.

No heat-engine which is in actual use works on the principle of a Carnot engine, and all actual heat-engines are quite frankly irreversible in Carnot's sense, since in none of them is the rise of temperature brought about by merely doing work on the working substance. In all of them the rise is mainly or solely produced by the direct application of heat. One reason why the Carnot principle is never adopted is that in the Carnot engine a large amount of unnecessary work is done on the flywheel during expansion, and by it during compression, the result being that for a given expansion, with its associated size of engine and friction, the net theoretical work in a cycle is small, and the horse-power also small, owing to the long time taken for a cycle with the piston travelling not too fast for satisfactory thermal efficiency. Losses from friction, and from leakage of heat of all kinds, are therefore excessive. The theoretical indicator diagram of a Carnot engine, whether the temperature difference is great or small, is similar to a thin crescent-moon; and this is a hopeless shape for the development of horse-power and avoidance of overwhelming loss from friction and heat leakage.

Whether in an actual Carnot engine, or any other form of heat-engine, it is not the relative excess in absolute temperature of the source of heat above that of the condenser, but the relative excess in absolute temperature of the working substance in expansion above its temperature during compression, and the rate of movement, that directly determine the theoretical thermal efficiency. In no actual heat-engine does the temperature of the working substance rise as high as that of the source. It is not the excess in temperature of the source, but that of the working substance, that is of importance.

In a gas-engine of the Carnot type, or any other type with external heating, rise in temperature of the working substance would in practice be limited, owing to the practical impossibility of communicating to the bottom of the cylinder or other immediate source of the heat a very high temperature; but this objection does not apply to a gas-engine with internal heating, like the internal combustion-engine.

Let us now consider another hypothetical form of heat-engine, in which the cylinder was surrounded by a water-jacket supplied with a stream of cold water, and in which a definite amount of heat was, commencing with the beginning of each expansion, communicated at a definite rapid rate by electrical means to the enclosed air or gas, so large an amount of heat being added as to produce a great rise of mean temperature in the gas during slow expansion. The mean temperature would rise to a maximum which, during slow enough expansion, would depend almost entirely on the relative rates of supply of heat and simultaneous absorption of it by the cooled walls. When the supply of heat was cut off the gas would cool itself rapidly by continued loss of heat to the walls, and to only a slighter extent by doing work during the further expansion. We may suppose the expansion continued only until the temperature fell to that of isothermal compression, and that the whole of the compression was, on an average, isothermal, just as in the first part of compression in a Carnot engine. During compression the walls, which had retained extra heat during expansion, would be giving off this extra heat, together with the relatively larger amount of heat formed during compression, to the water in the jacket. This engine differs from the Carnot engine inasmuch as not only is the temperature raised by nothing but the application of heat from an inside source, but a certain definite amount of heat is applied in each cycle, and during the whole of expansion heat is being lost to the outside, as well as during the whole of compression. The communication of heat is also far more rapid than in the Carnot engine.

The walls would evidently possess a mean temperature, just as in the Carnot engine; but this mean temperature would always be below that of the gas during isothermal compression. The temperature of the gas in isothermal compression would nevertheless be higher in virtue of extra heat communicated to the walls during isothermal expansion, as in the Carnot engine, so that in consequence of the more ready leakage of heat into the walls during expansion and less ready leakage into them during compression, the temperatures of expansion and compression would be brought closer together, just as in the Carnot engine.

As we increased the piston velocity from a very low value by reducing the external resistance, less and less of the heat communicated would be lost to the jacket in expansion and by friction. The mean temperature at which work was done in expansion would therefore rise, with corresponding increase in the work done in expansion. In other words, more of the heat supplied would be converted into mechanical energy, and less absorbed by the jacket, during expansion. At the same time, however, since the piston velocity during compression would be increased, more heat would pass into the jacket during compression, and the temperature of compression would rise.

At first this latter rise would be small, while the increased work done at the expense of heat previously going into the jacket would be large. With further increase in velocity less and less, but in diminishing proportion, of the heat supplied would have time to pass into the jacket during expansion, while more and more, but in diminishing proportion, would be converted into mechanical energy during expansion. But of what was so converted more and more would be reconverted into heat during compression, and pass then into the jacket. At a certain velocity the increasing amount passing into the jacket during compression, at the increasing temperature of the gas during compression, would balance the diminishing amount passing in during expansion, so that at this point, when the heat lost by leakage in expansion was equal to the mechanical energy reconverted into heat during compression, the net proportion converted into mechanical energy would cease to increase. The ratio of work done in expansion to work done in compression would depend on the mean absolute temperatures at which each kind of work was done.

With further increase in velocity the leakage of heat during expansion would further diminish, and the conversion of heat into mechanical energy during expansion would correspondingly increase, but only at the cost of a greater reconversion of mechanical energy into heat passing into the jacket during compression, since the reconversion varies as the square of the velocity, while the corresponding conversion varies at a less rapid rate, since the mass being set in motion is diminished. Thus further increase in velocity would diminish the efficiency, which would, at its maximum value, be the same as if the loss of heat to the jacket were just double the loss due to the heat formed in compression. The rate of work done by the engine would, however, at first increase with the increasing velocity.

The engine itself might, if it were not governed, suffer damage owing to its rate of running, if the external resistance was diminished too much, and each communication of heat was almost instantaneous. As, however, each communication of heat takes an appreciable time, this would itself limit the increase in velocity, since the heat communicated would not have time to disappear during the doing of work, and would thus remain in the gas and raise the temperature of compression, till the engine ceased to accelerate.

The engine under discussion is a contrast to the Carnot engine with its diminishing supply of heat per cycle as the velocity increases; and all practical heat-engines differ in a more or less similar manner from the Carnot engine.

Since, in the cycle of this engine, the temperature is raised by the application of external heat, we cannot calculate its efficiency from the actual upper and lower temperatures, as with the Carnot engine. As, however, the temperature of the whole of compression is isothermal or virtually so, we can imagine that it has also a virtual temperature of isothermal expansion, this temperature representing what would be the average temperature of work in expansion if the specific heat of the gas were constant. Since some of the work will be done at a much lower temperature than the maximum temperature which would exist if the specific heat of the working substance was constant, the virtual temperature of isothermal expansion will of course be well below the maximum temperature which would exist if the specific heat of the working substance was constant, but will be only slightly below the actual maximum temperature, since, owing to the increase in specific heat at high temperatures, more heat is communicated to the working substance and converted into work than corresponds to the actual rise in temperature.

If we call the virtual temperature of isothermal expansion T, and the virtual temperature of isothermal compression T_1, the maximum thermal efficiency, when the piston velocity has sufficiently increased to realise this, will be $\frac{T - T_1}{2T}$. With gradually increasing velocity T will at first rise proportionally more than T_1, until with further increase the diminished loss of heat to the jacket in expansion is balanced in its effect on $\frac{T - T_1}{T}$ by the increased loss of heat during compression. The maximal efficiency is $\frac{T - T_1}{2T}$, but, in contrast to the Carnot engine, the ideal engine now considered may attain to a much larger difference between T and T_1; and this is characteristic of all practical heat engines. Variations in the area of the indicator diagram are also in this and other possible heat-engines a reliable index of varying thermal efficiency, which, as was pointed out above, is not so for the Carnot engine.

At an extremely low velocity the indicator diagram would be scarcely more than the lines of isothermal expansion at the temperature of the water-jacket, with a high sharp-pointed spike, representing a "knock" in the engine, at the beginning of expansion; and the area enclosed in the spike would be very small. At this velocity both rise and fall of temperature in the cycle would reach their maximum values, but the thermal efficiency would have fallen to practically zero, since with any lower velocity the engine would stick at the end of the compression stroke, owing to friction. We can see from this example how absolutely mistaken it would be to imagine that the efficiency of a gas-engine can be anything even remotely like proportional to the maximum relative differences in absolute temperature which can be reached in its cycle.

In the engine under consideration the maximum difference between the virtual T and the virtual T_1, and, as a final result, the value of V, would depend on the amount of heat communicated, and its rate of communication, during the first part of each slow expansion, compared with the rate at which heat was lost by the gas to the walls. In contrast with the Carnot engine, we could, with this form of heat-engine, increase the rate and amount of heat communication without increasing at the same time either the temperature of the immediate source of heat or the cooling surface of the walls. The heat would, moreover, be going straight into the gas, without the delay involved in heating or cooling the bottom of the cylinder, or the consequent waste of heat which is more or less common to the Carnot engine and the original form of Newcomen engine before Watt's improvement of it. Owing to the effective cooling arrangement, the temperature of compression would at the same time be kept down. Thus the engine would work far faster, and, since the difference between T and T_1 could be increased to a far greater extent, with far greater thermal efficiency than the Carnot engine. At the same time the engine would be practically useless, since we can convert electrical into mechanical energy much more easily and economically than by first converting it into heat.

Let us now see how the efficiency would be affected by rise in the mean temperature of the walls. This could be effected by permitting the water in the jacket to become hotter. The result would be that the virtual isothermal temperatures of both compression and expansion would be raised; and whether the efficiency would be affected favourably or unfavourably would depend on whether the fraction $\frac{T - T_1}{T}$ was increased or diminished. With a low piston velocity T_1 would be raised to slightly above the raised mean temperature of the walls. Since, moreover, the gas would be hotter at the end of compression, the heat then applied would raise T correspondingly higher, the rate of heat-loss by the gas to the walls being the same, since their temperature is correspondingly raised. The result is that the efficiency would fall, since the value of $\frac{T - T_1}{T}$, or $1 - \frac{T_1}{T}$, would fall. At higher velocities this fall would still be present, since the value of $\frac{T - T_1}{T}$ at a higher velocity is dependent, as shown above, on its value at a low velocity. Thus the maximum efficiency of the engine would be reduced.

At first sight it might seem that by cooling the cylinder walls during expansion, in addition to the required cooling during compression, we are necessarily reducing the thermal efficiency by throwing away heat which might have been converted into mechanical energy but for the necessity of keeping the metal and lubricating oil sufficiently cool. This reasoning would, however, be erroneous. If there were no cooling during expansion the cylinder walls would soon become hotter, and would take on a higher temperature between the highest and the lowest temperatures in the cycle. As heat would always be leaking into or out of the walls in any case, the effect would be that though the temperature of the gas, and consequently the work done while the temperature of the gas was above that of the walls, would be increased by the absence of the cooling, there would be a corresponding increase in the rise of temperature of the gas over that of the condenser while the temperature of the gas was below that of the walls, with consequent loss in thermal efficiency.

It is only a moderate step from the engine just considered to the ordinary type of internal combustion-engine. The influences of the differences between the two types of engine may be considered one by one. The four-stroke or Otto type of cycle will be discussed, as there is no essential distinction between it and the two-stroke or Clerk cycle.

The first difference is that at the end of expansion the vitiated air is discharged in the succeeding stroke, and a corresponding volume of pure air, or pure air mixed with gas or vapour, is taken in during the next stroke. This air or mixture is then compressed in the following stroke. The mixture, if it was a mixture, is then ignited by a spark or other means. If the compressed gas was simply air, the fuel, in the form of finely sprayed oil or gas, is introduced at the end of compression, and then fired. In the Diesel type of engine the firing is due to the fact that some air has been so much heated in the compression that firing, even with heavy oil as the fuel, occurs at once. The combustion, in whatever way brought about, is not so rapid as to cause too sudden a rise of pressure, but sufficiently rapid to give a high thermal efficiency and horse-power. The heat is of course produced by the combustion of the fuel. If, owing to imperfect mixture, excess of the fuel, excessive speed of the engine, or other causes, some of the fuel is not burnt, or burnt too slowly, as very usually happens, the thermal efficiency must fall off correspondingly; but the accompanying loss of horse-power may be more than compensated for by the greater speed. If, on the other hand, there is too little fuel for the oxygen present in the air, the thermal efficiency, as well as the horse-power, must likewise fall off; and if the engine happens to be travelling fast enough at the time it will act as a brake, as in the case of a motor-car engine when the car is travelling fast and the concentration of petrol mixture per cycle is reduced.

The second difference is that in place of the heat formed during compression going wholly into the water of the jacket, the hot air left at the end of expansion is discharged to the outside, any excess heat contained in it, and that formed by friction as it escapes, going into the atmosphere. By listening to the sound of an unsilenced motor bicycle we can realise how great this latter heat is. Cool fresh air is then taken in, and heated by compression in contact with the walls, till it may reach what had been the initial temperature of the exhaust gas. As much heat is lost to the outside, the virtual temperature of isothermal compression becomes a very indirect conception.

The compression has sometimes been regarded as adiabatic; but this of course is impossible; and heat which would otherwise have passed into the jacket during compression is expended in heating the outside air, so that this air is sharing with the jacket the work of a condenser. The function of the water-jacket, or of direct air-cooling with a small cylinder, is occasionally referred to as being merely that of keeping the metal and lubricating oil cool. In actual fact the water-jacket is, or might be, the main condenser; and, as Carnot pointed out clearly, no heat-engine can work without a condenser. Without the jacket or special air-cooling the engine would become increasingly inefficient, and might become red-hot and give way, if it had not already been stopped by friction or seizing of the metal. It is evident that the water-jacket and the atmosphere act between them as the condenser. By limiting the rise of temperature in the walls of the cylinder, and consequently in the gas, the waste of heat to the air after the end of expansion is diminished greatly, and the undue rise in temperature which would otherwise occur in the gas during compression, owing to heat communicated by the walls to the gas, is also prevented. But for this latter effect the *relative* difference in virtual absolute temperature between the gas in expansion and in compression would be diminished, and the thermal efficiency would fall accordingly. The heat could still be got rid of in the exhaust air, so that but for the effects on the metal and lubricant the engine could still work in a cycle; but the exhaust gas would be very hot, and the effect on thermal efficiency would be the same as if the gas in the engine previously considered were allowed to become very hot during compression.

It is thus, from the direct standpoint of thermal efficiency, important to prevent, as far as possible, rise of temperature in the water coming from the water-jacket, and consequently in the exhaust gas, though of course it is usually not possible to prevent a considerable rise, and consequent diminution in thermal efficiency. The water-jacket is, in fact, a condenser, and not primarily a mere means of keeping the metal and oil cool in expansion. Heating of the atmosphere by the exhaust gas is only a supplementary means of condensation. For the sake, however, of increasing the rate of combustion it is necessary in these engines to let the virtual temperatures of expansion and compression rise, even though this permits increased combustion and waste by volatilisation of the lubricating oil.

When the water from the jacket and the exhaust gas are, perforce, allowed to become very hot, it is possible to obtain work from them; and this is done in the Still engine, where the waste heat is used to generate steam, which is applied to help the engine on. By this means a thermal efficiency as high as 38 per cent. has been obtained, though otherwise the maximal thermal efficiency seems to be only about 33 per cent. in the most economical internal combustion-engines at present in use. This efficiency could certainly be improved on if sufficiently rapid combustion could be obtained with a cool condenser.

In a small internal combustion-engine, other things being equal, the variation in temperature of the gas owing to leakage of heat into and out of the cylinder walls tends to be very rapid. Hence the frequency of cycle necessary is greatly increased, and thus becomes too much to permit effective or complete combustion of oil fuel. The combustion, if it has not been completed in the first part of the expansion stroke, seems to be slowed down owing to the rapid cooling effect of the great rate at which work is being done in the middle of the stroke.

With fuel in the much more expensive form of vapour or gas, however, the rate of combustion can be greatly increased, so that small petrol engines, as used for motor cars, aeroplanes, etc., have developed to a wonderful extent. They are capable of a good thermal efficiency, and give enormous power for their size, so that where cost of fuel does not matter they are at present unrivalled. In ordinary petrol engines, however, a large part of the fuel escapes combustion, with consequent drop in the thermal efficiency and emission of very poisonous gas.

In order to maintain a sufficient thermal efficiency and horse-power with varying resistances it is necessary to have changeable gear. Both the frequency of cycle and amount of work done per cycle can thus be kept approximately steady. We cannot greatly diminish the frequency or increase the heat supply per cycle, to suit increased resistance, without very seriously affecting the thermal efficiency, owing to excessive leakage of heat through the cylinder walls or ineffective combustion of fuel. With sufficient increase in resistance the engine is thus brought to a stand. On the other hand, if we sufficiently diminish the resistance, so that the cycle is much too frequent, and the virtual compression temperature is excessive, the total quantity of the work in compression comes at last so near to that in expansion, that the engine cannot run any faster. If a greater velocity is communicated to it from outside it will act as a brake, as of course it will do in any case if less fuel is supplied to it and it remains in motion and in gear.

By increasing the concentration of the combustible mixture during the effective part of the stroke, as in the Diesel engine with its high compression ratio, we increase the rate of heat-production, thus increasing the efficiency by obtaining a much more effective expansion. If, at the same time, the temperature of the gas is allowed to rise much higher before ignition, as in the Diesel engine, the rate of combustion is also increased. It would seem, however, that the Diesel principle of ignition is inconsistent with obtaining the maximum thermal efficiency which the engine might directly give, since the temperatures of the exhaust gas and water from the jacket are too high. Were this not the case it would be impossible to obtain a very definitely higher efficiency by utilising the surplus heat to produce steam, as in the Still engine.

It is sometimes imagined that the compression stroke of an internal combustion-engine is adiabatic. This is of course a mistake. During the stroke in which the cylinder is taking in fresh air, and the first part of compression, the walls are warming the air, while during the rest of compression they are cooling it to an increasing degree. It is only in part of the air that compression is nearly adiabatic.

In practical heat-engines it is usually not the maximal thermal efficiency, but the maximal horse-power for a given size or weight of engine that is important. In the engines of an aeroplane, for instance, the latter factor is of supreme importance, and in other engines is of at least considerable importance owing, if for no other reason, to diminished capital expenditure for a smaller size of engine. Now the greater the *absolute*, not merely relative, difference between the virtual temperatures at which work is done in compression and expansion respectively, the faster can a heat-engine run, and consequently the greater the horse-power, for a given thermal efficiency; but, in spite of ineffective combustion, the horse-power continues to increase with increased speed after the thermal efficiency has considerably diminished. Where size of engine is important it thus becomes important, and may be essential, to let the engine work at a considerably greater speed than is consistent with maximal thermal efficiency. An apparent reckless waste, in the engine of an aeroplane or even a motor-car, of heat capable of conversion into mechanical energy, or of potential heat as unconsumed fuel, can thus be justified in so far as the maximum horse-power is actually being obtained, which is by no means always the case. Rate of combustion of the fuel, dependent as it has been shown to be on turbulence and other factors, such as intimate mixture of fuel and air, is a matter of very great importance. To this matter much attention has been successfully given, and is still being given, by mechanical engineers.

The internal combustion engine possesses the great advantage over other types of engine that the source of heat is inside the engine, thus saving space and preventing direct loss of heat from the source to the outside. A very great rise in temperature is also produced in the cycle of an internal combustion engine with gas or oil as the fuel, and air along with the gaseous products of oxidation as the working substance. On the other hand, with a given length of stroke we can only vary within comparatively narrow limits the heat supply per cycle, since with either too little or too much fuel the fuel does not burn at all, or burns very imperfectly. An internal combustion engine is thus, in contrast to a steam-engine, comparatively inelastic in its capacity for encountering varying loads, though this disadvantage can easily be neutralised by the provision of changeable gear.

In another respect also it has marked limitations. These are dependent on the rate of combustion. As we make the cylinder of an engine smaller we increase the ratio of its surface to its capacity, since the former varies as the square, and the latter as the cube, of the diameter. We thus necessitate an increased rate of supply of heat, so that, if the engine is to work with the maximum efficiency it can attain, it must work faster. But if we attempt to work the small engine fast enough for maximum efficiency, we are faced by the trouble that owing to the time allowed for effective combustion being too short the value of h is diminished. At a very high altitude an internal combustion engine becomes almost useless unless a high compression ratio is provided for, to compensate for the diminished barometric pressure.

In a large internal combustion engine, and particularly with a high compression ratio, the value of V becomes so low that the piston cannot move very fast. Thus, though the combustion is very effective, the horse-power does not increase in proportion to the size of cylinder. On the other hand, the weight of metal needed to give the requisite strength greatly increases in proportion to the capacity. Hence large internal combustion cylinders are very heavy for the horse-power which they give, and the horse-power for a given weight can be much better increased by using small cylinders where extra cost of fuel is practicable.

It follows from the foregoing discussion that the thermal efficiency of an internal combustion engine has no direct relation to the very great relative differerence between the maximum absolute temperature attainable in expansion and the temperature of the outside air. What is important is the relative difference between the virtual temperatures of isothermal expansion and compression. If we call these temperatures T and T_1 the maximum efficiency is $\frac{T-T_1}{2\,T}$. In a Diesel engine running with its greatest thermal efficiency, T is much below the maximum temperature, and T_1, which is about the temperature at the end of expansion, is far above the air temperature. The value of $\frac{T-T_1}{T}$ seems to be about 75 per cent., as the brake efficiency is only 33 per cent. Frictional loss is small at the most favourable piston velocities. The indicator diagram exaggerates it (see p. 318).

Starting from principles laid down by Carnot, but developing these principles in the light of the dynamical theory of gases and temperature, we have been led to the conclusion that a heat-engine with a perfect gas as its working substance is incapable of converting into mechanical energy more than half of the heat energy which it could convert if reversible engines were possible. It is well known that no heat-engine has ever yet been known to work with as much as 50 per cent. efficiency; and Joule had already pointed out in 1843 that the best heat-engines then existing gave a yield corresponding to less than 10 per cent. of the mechanical equivalent of heat. At the present time the yield of the most economical heat-engines is not quite as much as 40 per cent.

The reasons for the low efficiency have, however, hitherto been obscured by the acceptance of reasoning based on Carnot's mistake in assuming that a reversible heat-engine is possible. The gas laws follow from the dynamical theory of gases and heat, and it appears that it also follows from this theory and its extensions to liquids and solids that in heat-engines working under the conditions which exist in their actual working not more than half the heat energy communicated to a perfect gas can, under the most favourable conditions conceivable, be converted into mechanical energy.

Let us endeavour to obtain a general view of why this is so. We have seen above that Carnot, though he was right in maintaining that no heat-engine could work more economically between two temperatures in converting heat into mechanical energy than if it were working reversibly, was wrong in supposing, as he did, that a reversible heat-engine is a possibility. Nor is it true that if such an engine were a possibility it would necessarily be any more economical of heat than a heat-engine of the ordinary type in which the part of the cycle in which the same amount of heat is applied is completely irreversible. This can easily be shown to be untrue unless we not only make the impossible condition that one of the engines is reversible, but add the condition that the engines work between the same extremes of temperature.

If Carnot had been right in assuming that while net work was being done the temperatures of the gas during expansion and condensation could be equal to those of the source of heat and condenser respectively, we might have said that the efficiency of his imaginary reversible engine would vary as $\frac{T - T_1}{T}$ when T was the absolute temperature of the source of heat, and T_1 that of the condenser. By sufficiently raising, as can easily be done, the value of T, with the condenser at atmospheric temperature, we could have thus quite easily obtained an efficiency approaching 100 per cent.

As we have seen, we could only do this if, firstly, the engine were moving very slowly, and, secondly, if we could prevent loss of heat from occurring in any other way than through conversion into mechanical energy. As regards the second condition, it is evident that the more slowly the engine moves the greater will be the proportion of heat which escapes per cycle if such escape occurs, which it undoubtedly must, in accordance with the dynamical theory of heat. We are led away from actuality by the convenient words "adiabatic" or "isentropic." Actually it is impossible to prevent escape of heat, though with great velocities we can reduce it to a minimum, as in the propagation of a sound-wave. This is because the substances of which the engine is made consist of atoms and molecules which take up and transmit the chaotic energy which we call heat. Thus heat is leaking away all the time the gas in the cylinder is above the surrounding temperature; and even if we surrounded the cylinder with a vacuous space, and minimised radiation, the leakage would continue into the walls, and finally into the condenser. In order to restrict as far as possible the effect of the leakage on thermal efficiency we must let the engine work as fast as possible without thereby diminishing too far its actual net thermal efficiency owing to the piston velocity being too great; and the rate which gives maximum efficiency gives only an efficiency never reaching 50 per cent. Unless the engine is working fast enough, its thermal efficiency, and still more its horse-power, is so small as to make it useless. It would, however, often have been convenient if, by running an engine slowly, we could also have run it with as great or greater thermal efficiency, and therefore more economically, just as we can run a mill-wheel more economically of water if we run it slowly.

The reason why more of the heat energy is not available is that the movements of the molecules, whether in the gas or in the walls of the heat-engine, are chaotic. If this were not the case a heated gas, as we have seen, might lose all its imparted heat energy by doing work in expansion, and then be returned quite rapidly to its original state without appreciably more heat being re-imparted to it than if it were recompressed very slowly. We can easily convert directed forms of energy, whether of molecules or electrons, and whether kinetic or potential, into mechanical energy with nearly a hundred per cent. efficiency; but we cannot do this with the chaotic form of molecular energy which we call heat; hence we cannot, by merely reversing a process by which heat at a high temperature has been converted into mechanical energy, return the heat to its high temperature without great loss of energy.

Since this follows from the dynamical theory of heat, and since heat is constantly being produced by the interaction of bodies, it also follows that if the visible universe is a chaotic mechanical universe, such as the Newtonian physics, including the dynamical theory of gases and the other states of matter, really assumed it to be, its energy is constantly running down irreversibly into heat at an even temperature. There is no need, however, to postulate this as a mere empirical law under the name of the "Second Law of Thermodynamics"; and indeed this so-called law, as ordinarily stated, is misleading, since, as so stated, it is consistent with the existence of reversible heat-engines. In so far as we accept the dynamical theory of gases and the other states of matter, and at the same time admit, as we must, the fallacy of Carnot's assumption that a heat-engine may be reversible, all the truth, or relative truth, that is summed up in the Second Law of Thermodynamics as applied to gases becomes simply a part of ordinary Newtonian physics as extended to molecules. The question whether the energy of the universe is running down into heat becomes, then, a part of the wider question whether the Newtonian assumption is correct that the visible universe is mechanical in its nature. Those who regard this as being, ultimately considered, a mere subjective assumption inconsistent with a great part of our experience, even apart from experience of life and conscious behaviour, may nevertheless also regard it as being for certain immediately practical purposes a most useful principle. It must also be remembered that our very conceptions of heat, temperature, the states of matter, and matter and energy themselves, are based on the Newtonian assumption and are thus of only provisional value. Molecular or intramolecular activity which is not chaotic cannot be regarded as heat. Nor could what was not a chaotic system be regarded as a gas.

The failure on the part of Kelvin, Clausius, and those who have followed them, to recognise the implications of the fact that heat energy is chaotic has made the existing academic account of heat-engines extremely misleading to mechanical engineers. It has held out to them illusory prospects of very high thermal efficiencies, given rise to confusion in the actual design and improvement of heat-engines, and, as we shall see, has very seriously retarded the development of the steam-engine. It seems to me impossible to disguise this failure, however sincerely we continue to honour, for their other actual achievements, those who failed.

The subject of thermodynamics has been discussed in the present chapter in so far as deductions as to the conversion of heat into mechanical energy can be drawn from the dynamical theory of a perfect gas. A further discussion, including in its scope liquids, and referring more particularly to the steam-engine, must be postponed till Chapter VIII., after the connection between the gaseous and liquid states has been discussed.

CHAPTER III

GASEOUS DIFFUSION AND OSMOSIS

The subject of gaseous diffusion and osmosis will be considered in the present chapter. In the application of the dynamical theory to liquids, and to the facts of physiology, diffusion and osmosis become, as will appear in later chapters, subjects of fundamental importance.

When two different sorts of gas are brought into connection with one another, a mutual interpenetration of the molecules must, on the dynamical theory, occur unless, as is frequently the case, there are special conditions which prevent this interpenetration. The process of interpenetration molecule by molecule is known as diffusion. It must occur without any loss of kinetic energy whatsoever by "friction," and apart from the existence of special conditions its occurrence depends simply on the probabilities as to the ultimate mean distribution of different kinds of molecules, free, except for hindrances due to molecular collisions, to move about in the space within which the substances are confined.

Owing to the existence of membranes which are permeable to one sort of molecule, but not, or only to a very slight extent, to another sort, we can, so to speak, suspend in one respect the abstract law of probability, and thus subject the factors in the process of diffusion to experimental analysis. By means of the dynamical theory we can also deduce the laws of gaseous diffusion; and as already mentioned, one of these laws—that of Graham—was so deduced by Waterston in the case of gases. It was Graham who also first discovered the existence of membranes which possess the property of being permeable to one substance, but not, or only to a slight extent, to others. Thus he found that hot palladium or platinum is permeable to hydrogen, but not to other gases, while india-rubber is far more permeable to carbon dioxide than to air or even hydrogen, and considerably less permeable to nitrogen than to oxygen.[1] He also distinguished "colloids" as substances to which, when in solution, vegetable parchment membranes are impermeable.

Let us imagine pure hydrogen and pure nitrogen at ordinary atmospheric pressure and contained in two equal gas-tight chambers separated from one another by a rigid septum permeable to the hydrogen but completely impermeable to the nitrogen. The hydrogen contained in chamber 1 will immediately begin to diffuse into the nitrogen in chamber 2, and will continue to do so until the pressure of the hydrogen is the same in the two chambers. The pressure in the first chamber (containing only hydrogen) will then be half an atmosphere, and that in the second chamber 1.5 atmospheres. If, however, the pressure in the first chamber is kept constant by reducing its volume or letting in hydrogen as required, the pressure in the second chamber will be two atmospheres. In any case, the extra pressure produced in the second chamber by the diffusion of hydrogen into it will be equal to the pressure in the first chamber.

---

[1] Graham, *Phil. Trans. Royal Society*, 1861, p. 183, and 1866, p. 399. Further discoveries in this direction will be discussed in Chapter IV.

Now let us suppose that instead of there being pure gas in the second chamber to start with, there had been a mixture of equal parts by volume of nitrogen and hydrogen, and that the pressure in this chamber was so raised as to stop diffusion into it. The pressure required would evidently be two atmospheres, or one atmosphere in excess of the orginal pressure. But the original difference in pressure of the hydrogen between the two chambers was only half an atmosphere, or half as much. If, however, keeping the pressure in the first chamber constant, we allowed hydrogen to diffuse into the mixture until diffusion ceased owing to the pressure of the hydrogen being the same in the two chambers, the pressure of the mixture would only be one and a half atmospheres. In either case, we obtain a measure of the original difference in pressure of the hydrogen in the two chambers. By the first method the extra pressure indicating the difference is one atmosphere, half of which is due to the hydrogen, so that the actual original difference in pressure of the hydrogen is half an atmosphere. By the second method the extra pressure is half an atmosphere, all of which is due to pressure of hydrogen, so that the actual original difference is again half an atmosphere.

The first method would evidently be inapplicable if only pure nitrogen were originally present in the second chamber: for no measurable amount of external pressure would keep out the hydrogen. By the first method the mass of gas in the second chamber is kept constant, while by the second the volume of the second chamber is kept constant, and hydrogen allowed to diffuse in, till diffusion inwards is balanced by diffusion outwards; hence the extra pressure developed is a direct measure of the deficiency of pressure of the hydrogen in the second chamber.

The pressure which has to be applied to the mixture in order to stop diffusion into it corresponds, as will be shown in Chapter IV., to the true osmotic pressure of a liquid. If we call this extra pressure P_o, the original atmospheric pressure in the two chambers P, the number of moles per litre of pure hydrogen N, of the mixture N_1, and of the nitrogen in the mixture n, so that the number for the hydrogen in the mixture is $N_1 - n$, we can represent the diffusion equilibrium by the equation $\frac{N_1 - n}{N} P + \frac{N_1 - n}{N} P_o = P$. Hence, since by Avogadro's law $N = N_1$, $P_o = \frac{n}{N_1 - n} P$. In other words, the osmotic pressure is equal to the original pressure multiplied by the ratio of moles of the gas to which the septum is impermeable to moles of the gas to which it is permeable. Since at low pressures $P = N \, .082 \, T$, we can also state the equation as $P_o = \frac{nN}{N_1 - n} .082 \, T$.

As will appear later, the failure of van't Hoff and his successors to realise that osmotic pressure depends upon the ratio of moles of solute to moles of solvent has been, and is, a source of endless confusion in physical chemistry. Osmotic pressure varies, not as moles of solute per litre, as van't Hoff thought was the case for, at any rate, dilute solutions, but as the ratio of moles of solute to moles of solvent.

In the particular case of any ordinary gas at a pressure of one atmosphere or less, the values of N and N_1 would be practically identical. We could thus write the expression for osmotic pressure as $\frac{nN}{N-n}.082\,T$; but the substitution of N_1 for N is more explicit.

Now $\frac{nN}{N_1-n}$ means, in the case before us, the ratio of moles of nitrogen to moles of hydrogen present in the mixture multiplied by the number of moles of hydrogen in one litre of pure hydrogen at the same pressure and temperature. If S is specific gravity of hydrogen at existing pressure and temperature, referred to water at 4°, $\frac{nN}{(N_1-n)S}$ represents moles of nitrogen per 1000 grammes of hydrogen multiplied similarly. Hence the osmotic pressure is given at low pressures by multiplying the moles of nitrogen per 1000 grammes of hydrogen by .082 T (or RT), and by the moles per litre of the pure hydrogen. A similar calculation, as will appear in Chapter IV., gives correctly the osmotic pressures of solutions, as well as their depressions of freezing-points and elevations of boiling-points.

If, without varying the pressure, we raise the temperature of the gases in the two chambers, the value of the osmotic pressure will not be altered, since by Charles's law the value of N will vary inversely as that of T. If, however, the temperature were increased without allowing the gas to expand, the osmotic pressure would increase in proportion to the increase in T, since the value of N would not alter, but the value of P would increase in the same proportion as T.

The expression $\frac{n}{N_1 - n}$ P, when applied to a gas at only about or below atmospheric pressure, represents, as explained in Chapter I., not only pressure, but the energy in litre-atmospheres required to compress one litre of the mixture at constant temperature to such an extent that the pressure of the hydrogen in it is equal to that of the pure hydrogen. But $\frac{n}{N_1 - n}$ P litre-atmospheres $= \frac{nN}{N_1 - n}$ .082 T, or $\frac{nN}{N_1 - n}$ 1.983 T calories. The great importance of this relation will appear in Chapter V.

If only pure nitrogen were originally present in chamber 1, no definable amount of added pressure would suffice to keep out the hydrogen. The osmotic pressure would thus rise to infinity. On the other hand, if only 1 per cent. by volume of nitrogen was present along with 99 per cent. of hydrogen, the osmotic pressure would be only 1 per cent. greater than the original partial pressure of the nitrogen; and with still lower percentages the difference between the osmotic pressure and the partial pressure of the nitrogen would tend to become inappreciable. In this connection it must be borne in mind, however, that, as already shown, in a gas mixture at great molecular concentration the partial pressures of individual gases with different sizes of molecules are not proportional to the numbers of their molecules present in a litre.

It is evident that osmotic pressure, as defined above, is identical neither with the original partial pressure of the gas to which the septum is impermeable, nor with the original deficiency in the partial pressure of the gas to which the septum is permeable, as compared with its pressure in the pure gas. On the other hand, the cause of the diffusion of gas from one chamber to the other is simply the fact that in one chamber the partial pressure of hydrogen is higher than in the other. The initial difference of pressure is evidently $\frac{n}{N}$ P, since the pressure of the hydrogen on one side is $\frac{N_1 - n}{N}$ P, and on the other P, and in the case of a gas at atmospheric pressure N_1 and N may be taken as equal, so that $\frac{N_1}{N} P = P$. But as pressure must be applied in proportion to the ratio by volume between the constituents of the mixture in order to make up by externally applied pressure for the deficiency in the pressure of the hydrogen in the mixture, the pressure which must be applied is greater than the arithmetical deficiency in hydrogen pressure, and becomes enormously greater when the percentage of hydrogen is very low. If, however, we stop the diffusion, not by increasing the pressure of the mixture, but by depressing the pressure of the pure hydrogen, the required depression (P_d) is not $\frac{n}{N_1 - n}$ P, but only $\frac{n}{N}$ P, or $\frac{n}{N}$ N .082 T, which equals n .082 T, or nRT. The pressure to be taken off for each arithmetical increase in the proportion of nitrogen is thus proportional to the arithmetical increase.

Now let us see what must happen when the two gases present are at very high pressure. At high pressures it becomes necessary to draw a sharp and clear distinction between the pressure (bombardment pressure) of a gas and its diffusion pressure into another gas. In diffusion the molecules of one gas penetrate into the virtual free space between the molecules of another gas; and the diffusion pressure of one gas into another is simply the initial bombardment pressure produced per unit of sectional area by the molecules which are able to penetrate into the second gas when the molecules of the latter gas are prevented from leaving a space within which they are confined. The condition that one gas can diffuse in while the other is prevented from diffusing out is fulfilled through the interposition of a semi-permeable septum between two otherwise gas-tight chambers, as already described. Of the molecules of a gas which penetrates into a second gas by diffusion the number present in the second gas will, in accordance with the law of probability, be unable to increase as soon as diffusion pressure backwards balances diffusion pressure forwards.

The proportion of free space available for penetration is given by the value of $1 - N_1 b_1$ for the gas which is being penetrated. As soon as the concentration of the penetrating gas in this free space is equal to its concentration in the space containing the pure penetrating gas, diffusion backwards must equal diffusion forwards, so that the pressure in the penetrated gas will cease to increase. Since, moreover, the extra pressure is only applied to the portion $1 - N_1 b_1$ of the penetrated gas the initial diffusion pressure only amounts to $(1 - N_1 b_1)P$ when P represents the bombardment pressure of the pure penetrating gas.

In connection with this reasoning, it must be quite clearly borne in mind that on the dynamical theory gas is not a homogeneous fluid, nor is there any such thing in nature as a homogeneous fluid. We may for purposes of practical convenience in dealing with gases or liquids represent them to ourselves as if they were continuous homogeneous fluids. But this is only a practical rule-of-thumb makeshift, quite inconsistent with the numerous facts on which the dynamical theory of matter is based. If gases were continuous homogeneous fluids the phenomena of osmotic pressure, along with endless other phenomena, would be completely unintelligible.

In a gas at ordinary atmospheric pressure the value of $1 - N_1 b_1$ is so very near unity that the difference may be neglected. Hence the diffusion pressure may be taken as identical with the full bombardment pressure of the diffusing gas. But at very high pressures $1 - N_1 b_1$ may be only a small fraction of unity, so that the diffusion pressure is likewise only a small fraction of the bombardment pressure of the pure diffusing gas.

We can now consider the case of gases at very high pressure in the two chambers with the semi-permeable septum between them. Let us suppose, to start with, that there is pure hydrogen in the first chamber, and pure nitrogen in the second, and that the pressure in the first chamber is maintained at the original value P by adding hydrogen as required. The initial diffusion pressure of the hydrogen into the nitrogen will evidently be $(1 - N_a b_a)$ P, when $1 - N_a b_a$ represents the proportion of virtual free space in the nitrogen. This, then, is the pressure of hydrogen which would be required to balance the diffusion of hydrogen from the pure hydrogen. But $(1 - N_a b_a)$ P is less, and may be far less, than P. Moreover, $1 - N_a b_a$ is, as explained above (p. 25), to $1 - Nb$ (the proportion of virtual free space in the pure hydrogen) as N_a to N, where N_a and N represent the number of moles per litre of pure nitrogen and of hydrogen. Hence, $1 - N_a b_a = \frac{N_a}{N}(1 - Nb)$. Hence the initial diffusion pressure of the hydrogen into the nitrogen is $\frac{N_a}{N}(1-Nb)P$. But $P(1-Nb)$, or $(1-Nb)P = N\ .082\ T$. Therefore the initial diffusion pressure is $\frac{N_aN}{N}\ .082\ T$, or $N_a\ .082\ T$, or N_aRT.

The expression $\frac{N_a}{N}(1-Nb)P$, or $N_a\ .082\ T$, represents also the translational energy, in litre-atmosphere units, required for the expansion of a gas against pressure P from the volume Nb to one litre. It is evident that this energy is less, and may be far less, than that required for expanding the gas from an initial volume so small that it can be disregarded.

Let us now suppose that the second chamber contains, not pure nitrogen, but a mixture of nitrogen and hydrogen. The initial diffusion pressure of the pure hydrogen through the septum into the mixture is $\frac{N_1}{N}(1-Nb)P$, and of the hydrogen in the mixture into the pure hydrogen $\frac{N_1-n}{N}(1-Nb)P$. The diffusion pressure of the pure hydrogen is therefore in excess by $\frac{n}{N}(1-Nb)P$. But $(1-Nb)P = N\ .082\ T$. Hence the excess is $\frac{nN}{N}\ .082\ T$, or $n\ .082\ T$, or nRT when units of pressure and volume are not specified.

Now $n\ .082\ T$, or nRT, is exactly the bombardment energy which the nitrogen would have if it were not only present alone, but if the equation $PV = .082\ T$ (or with the new notation $P = N\ .082\ T$) held good, which of course it does not: in other words, if Boyle's and Charles's laws in their original crude form held good for the nitrogen in spite of the great pressure. Of course, however, it is not PV (or with the new notation P), but $P(V-b)$ (or with the new notation $(1-Nb)P$) that equals .082 T (or N .082 T), at any pressure or temperature. As will be seen in Chapter IV., van't Hoff was misled through failure to realise this distinction, in addition to his failure to see that osmotic pressure varies as moles of a gas which cannot pass the septum to moles of gas which can. Osmotic pressure is not full bombardment pressure, but only that part of it (and at very high pressures only a very small part) which is effective in producing diffusion.

From the discussion at page 25 it is evident that $\frac{n}{N}(1-Nb)P$, or n .082T, does not represent the diffusion pressure (if it could manifest itself through the septum) of the nitrogen into the pure hydrogen. Its diffusion pressure would be $(N-N_1+n)$ .082T; and this, of course, in the particular case of a large-moleculed gas, would be greater than n .082T, since N_1 would be less than N, and at great pressures would be very much less. Endless confusion will be avoided in connection with diffusion pressure and osmotic pressure if this is borne in mind.

It will be noticed that if the pressure is very high the value of $1-Nb$ becomes much less than 1, and the excess of diffusion pressure of the pure hydrogen becomes correspondingly smaller as compared with its full bombardment pressure. If, for instance, the pressure is so high that the value of $1 - Nb$ is reduced to 0.1, the excess in diffusion pressure of the pure hydrogen will be reduced to a tenth of the excess in bombardment pressure. Hence, if we stop the diffusion of the pure hydrogen by increasing the pressure of the mixture, we only have to increase the pressure of the mixture to a tenth of what would have been necessary if the value of $1 - Nb$ in the mixture had remained at 1, as in a gas at atmospheric pressure.

We have now to find an expression for the amount P_0 by which the pressure in the mixture must be raised in order to stop the diffusion into it of the pure hydrogen. It has already been shown that at ordinary atmospheric pressure, where $1 - Nb$ may be taken as 1, and N_1 may be taken as equal to N, $P_0 = \frac{n}{N_1 - n} P$, or $\frac{nN}{N_1 - n} .082\,T$. At high pressures, however, N_1 may be very different from N, and $1 - Nb$ may be far less than 1, so that $(1 - Nb) P$ is far less than P. As already shown, when the two total pressures are equal the diffusion pressure of the pure hydrogen into the mixture is $\frac{N_1}{N}(1 - Nb) P$, and of the hydrogen in the mixture into the pure hydrogen $\frac{N_1 - n}{N}(1 - Nb) P$. When the pressure of the mixture is raised in order to equalise the two diffusion pressures, N_1 will become N_2; but the value of $\frac{N_1 - n}{N}$ will not alter, since each term in it will be increased in the same proportion. The value of $1 - \mathrm{N}b$ in the expression for the diffusion pressure of the mixture will also be reduced to $(1 - \mathrm{N}b) \times \frac{\mathrm{P}}{\mathrm{P}+\mathrm{P}_0}$. When diffusion equilibrium is established the following equation must therefore be satisfied

$$\frac{\mathrm{N}_1 - n}{\mathrm{N}}(1 - \mathrm{N}b)\mathrm{P} + \frac{\mathrm{N}_1 - n}{\mathrm{N}}\mathrm{P}_0 = \frac{\mathrm{N}_2}{\mathrm{N}}(1 - \mathrm{N}b)\frac{\mathrm{P}}{\mathrm{P}+\mathrm{P}_0} \times \mathrm{P}.$$

At comparatively low pressures, where there is a large relative difference between P and $\mathrm{P}+\mathrm{P}_0$, the value of $\mathrm{N}_2 \times \frac{\mathrm{P}}{\mathrm{P}+\mathrm{P}_0}$ is practically equal to N_1, in accordance with Boyle's law. If $\mathrm{N}_2 \times \frac{\mathrm{P}}{\mathrm{P}+\mathrm{P}_0}$ is taken to be N_1, the value of P_0, when the equation is solved, is $\frac{n}{\mathrm{N}_1 - n}(1 - \mathrm{N}b)\,\mathrm{P}$. At very high pressures, on the other hand, the relative difference between P and $\mathrm{P}+\mathrm{P}_0$ becomes very much smaller, since, as already shown, P_0 increases far more slowly than P. The gas is also far less compressible. Here again, therefore, though for a different reason, the value of $\mathrm{N}_2 \times \frac{\mathrm{P}}{\mathrm{P}+\mathrm{P}_0}$ is almost equal to that of N_1, without taking molecular compressibility into account. Hence even at high pressures the same expression would hold good approximately. To look forward to the next chapter, the osmotic pressures hitherto measured are only about 1 per cent. of the estimated internal bombardment pressure of water. Hence the maximum error of calculation from taking N_2 as equal to N_1

would be 1 per cent. We have still, however, to take molecular compressibility into account, and it will be shown in Chapter VII. (p. 253) that, even when molecules reach the concentrations existing in the liquid form, $N_2 \times \frac{P}{P + P_0}$ equals N_1, since molecules are compressible to just the extent needed to produce equality.

We may thus provisionally take the osmotic pressure as being exactly $\frac{n}{N_1 - n}(1 - Nb)P$, which equals $\frac{nN}{N_1 - n} .082\,T$. This means that at the same temperature the osmotic pressure varies as the moles of nitrogen (the non-penetrating gas) to moles of hydrogen in the mixture, or (as already shown) as the ratio of moles of nitrogen to 1000 grammes of hydrogen multiplied by its specific gravity (at the same pressure and temperature) referred to that of water at 4° C. This is a conclusion of great importance, as will appear in Chapter IV.

Let us now also consider what would be the extra pressure developed in the mixture if, keeping the volume of the second chamber constant, diffusion of the hydrogen into it at an initial pressure equal to the enormous internal pressure in the liquid state were allowed to continue until diffusion equilibrium was reached. As the virtual free space in the mixture would be very small—perhaps only a tenth of the gross volume—it is evident that only a very slight proportional increase in molar concentration would suffice to cause a very great increase in the pressure of the mixture. On the other hand, as has just been shown, the increased pressure required in order to balance the inward diffusion pressure would be only a small fraction of the excess in the bombardment pressure of the pure hydrogen over the bombardment pressure of the hydrogen in the mixture. Hence the pressure ultimately developed in consequence of diffusion inwards would be within a minute fraction the same as the true osmotic pressure as defined above, and the composition of the mixture would hardly be appreciably altered at any osmotic pressures hitherto measured in the case of liquids. Assuming, however, that the molecules are compressed to the extent assumed above, a further extremely minute amount of gas would enter; but this would be unaccompanied, up to high osmotic pressures, by any diminution in the total virtual free space in the mixture. Hence the mixture would behave, as regards diffusion into it, practically as if pressure were, from the first, externally applied to the extent required to balance the diffusion.

We can thus infer that for a mixture at very high pressure it becomes practically indifferent whether the osmotic pressure is measured by determining the pressure required to prevent diffusion, or by measuring the pressure produced when diffusion is allowed to occur, but the volume kept constant. As will be seen in Chapter IV., both of these methods have been used for determining osmotic pressures in the case of solutions, and they give indistinguishable results, though this would not be the case for gases at low pressures.

We can also equalise the diffusion pressures on the two sides of the septum by reducing the pressure of the pure hydrogen. In this case it is only to hydrogen, and not to another gas as well, that the change of pressure is applied. Hence the reduction is less (and when much nitrogen is present very much less) than the increase of pressure which has to be applied to the mixture. When equalisation is produced the diffusion pressure of the pure hydrogen into the mixture is, if we denote by P_d the depression of the hydrogen pressure, $(1 - N_1 b_1) P - P_d$; and the diffusion pressure of the hydrogen in the mixture is, denoting the reduced values in the cases of b and N as b_b and N_b, $\frac{N_1 - n}{N}(1 - N_b b_b)P$. But $1 - N_1 b_1 = \frac{N_1}{N}(1 - Nb)$; while $1 - N_b b_b = \frac{N_b}{N}(1 - Nb)\frac{P}{P + P_o}$. Hence we have the equation

$$\frac{N_1}{N}(1 - Nb)P - \frac{N_1}{N}P_d = \frac{N_1 - n}{N}(1 - Nb)\frac{N_b}{N} \times \frac{P}{P + P_d} \times P.$$

When, then, taking from Chapter VII. the same conclusion as before, we treat $N_b \times \frac{P}{P + P_d}$ as equal to N, it follows from the equation that $P_d = \frac{n}{N_1}(1 - Nb)P$. This equals $\frac{nN}{N_1}$ .082 T, or $\frac{nN}{N_1}$ RT. Hence, with N and T constant, P_d varies as $\frac{n}{N_1}$, or as the ratio of moles of nitrogen to total moles in the mixture.

Let us now see what will be the effect of raising the temperature while a very high pressure P is kept up. It is evident that since at very high pressures $1 - Nb$ is only a small fraction of unity a gas will expand only very slightly with a given rise of T. In other words, N will only diminish slightly with rise of T; whereas when P is low N diminishes, as we have already seen, in direct proportion to rise of T. It will be seen that as a consequence of this the osmotic pressure and osmotic depression, which, as already pointed out, are practically unaffected at low pressures by rise in temperature with constant pressure, are greatly affected at high pressures, so that with sufficiently high total pressures the osmotic pressure must increase nearly in direct proportion to the absolute temperature, though actually in proportion to NT. The volume occupied by a given number of moles must also be nearly constant at a very high pressure, though it actually varies slightly with T. The importance of these conclusions in connection with liquids will appear in the following chapters.

The above reasoning has been applied to a mixture of hydrogen and nitrogen, but is of general application in cases where the septum is permeable to one gas in a mixture, but not to the other gas. We can then call n the number of moles per litre of the gas which cannot pass the septum, and N the moles per litre of the other gas in the pure state. Hence P_o will then be equal to $\frac{n}{N_1 - n}(1 - Nb)P$, and P_d to $\frac{n}{N_1}(1 - Nb)P$.

As will appear in Chapters IV. and V., these conclusions are of far-reaching importance. The pressure $\frac{n}{N_1 - n}(1 - Nb)P$, or $\frac{nN}{N_1 - n}.082\,T$, or $\frac{nN}{N_1 - n}RT$, is the pressure which must be added in order to equalise the diffusion of the pure gas into the mixture. It is evidently quite erroneous to regard the mixture as possessing at its original pressure a positive osmotic pressure; and the confusion which has arisen from this error will be made evident in succeeding chapters. The osmotic pressure is the pressure which requires to be added in order to balance the diffusion of the pure gas into the mixture. We can speak quite correctly of the osmotic depression or deficiency of the gas mixture at its original pressure, but not of its osmotic pressure. The pressure $\frac{n}{N_1}(1 - Nb)P$, or $\frac{nN}{N_1}.082\,T$, or $\frac{nN}{N_1}RT$, is the osmotic deficiency of the mixture as compared with the osmotic pressure for the pure penetrating gas.

If we have the same gas mixtures in the two chambers, it is evident that when the pressures and temperatures are equal the gas which penetrates must be in diffusion equilibrium on the two sides, and that its diffusion pressure through the membrane will be $\frac{N_1 - n}{N}(1 - N_1 b_1)P$. If, however, without altering the pressure, we substitute for the non-penetrating gas an equal volume of some other non-penetrating gas with either larger or smaller molecules, we shall, if the pressure is very high, have either diminished or increased considerably the number of moles per litre in the mixture. As a consequence the value of $1 - N_1 b_1$ will be changed to $1 - N_2 b_2$, n to n_a, and N_1 to N_2. The diffusion pressure of the penetrating gas from the new mixture will, however, be the same, since the percentage by volume of the penetrating gas is the same, and consequently $N_2 - n_a$ is equal to $N_1 - n$. But the diffusion pressure of the pure penetrating gas into the new mixture will be different, since $1 - N_2 b_2$ equals $\frac{N_2}{N}(1 - Nb)$, while $1 - N_1 b_1$ equals $\frac{N_1}{N}(1 - Nb)$. The two diffusion pressures of the penetrating gas are thus $\frac{N_1 - n}{N} \times \frac{N_2}{N}(1 - Nb)P$, and $\frac{N_2 - n_a}{N} \times \frac{N_1}{N}(1 - Nb)P$, and it is clear that $(N_1 - n)N_2$ must be equal to $(N_2 - n_a)N_1$ in order that the two diffusion pressures of the penetrating gas may be equal. But if they are equal $\frac{N_1 - n}{N_1}$ must be equal to $\frac{N_2 - n_a}{N_2}$. In other words, the ratio of moles of penetrating gas to total moles must be equal in the two gas mixtures. It is thus evident that a gas in a mixture of gases is only in diffusion equilibrium with, or isotonic with, the same gas in another mixture at the same pressure and temperature if the ratio of its molecules to the total molecules in the mixture is the same. Since, moreover, $\frac{N_1 - n}{N_1}$ and $\frac{N_2 - n_a}{N_2}$ are respectively equal to $1 - \frac{n}{N_1}$ and

$1 - \frac{n_a}{N_2}$, it is equally evident that $\frac{n}{N_1}$ must equal $\frac{n_a}{N_2}$ if diffusion equilibrium exists—in other words, if the mixtures are isotonic.

In the above reasoning, the letters n and n_a have been used to denote the molar concentrations of the non-penetrating gases, so that $N_1 - n$ and $N_2 - n_a$ have denoted the molar concentrations of the penetrating gas. The reason for adopting this particular notation is simply that, as will be seen in Chapter IV., a similar notation has been generally adopted in connection with the osmotic pressures of liquids. If, however, we take n, n_a, n_b, etc., as representing the molar concentrations of any particular gases to which a septum is *permeable* in the cases of different gas mixtures, we can express more simply the conclusion just reached, since these mixtures will be in diffusion equilibrium as regards the penetrating gas when $\frac{n}{N_1}$, $\frac{n_a}{N_2}$, $\frac{n_b}{N_3}$, etc., are equal; and the equilibrium diffusion pressures of the penetrating gas will vary, at equal temperatures and pressures, as $\frac{n}{N_1}$, $\frac{n_a}{N_2}$, $\frac{n_b}{N_3}$, etc. Hence the equilibrium diffusion pressure of any constituent of a gas mixture into another mixture at the same pressure and temperature varies, not as the molar concentration of the constituent, but as the ratio of its molar concentration to the total molar concentration of the mixture. This conclusion, when it is extended to liquids, is of very wide-reaching importance in physical chemistry and physiology. It is not, for instance, simply the molar concentrations of water, or CO_2, or oxygen, that are of importance in connection with physiological reactions, but their diffusion pressures; and the same is true in connection with chemical equilibria in general.

The subject of diffusion and osmosis in liquids will be pursued in Chapter IV.

CHAPTER IV

OSMOSIS IN LIQUIDS

When two miscible liquids of different compositions are separated by a membrane of bladder, parchment, or any porous septum, or simply by gravity, an interchange, independent of gravitation, takes place between them, and liquid passes from one side to the other of the septum or dividing line. Thus pure water will pass into blood, or alcohol, or a solution of sugar or salt, against a pressure of several inches of liquid.

The vegetable physiologist Dutrochet, who investigated these phenomena about 1830,[1] called the flow into the liquid which increases in volume "endosmose"—literally, a thrusting inwards (from ἔνδον and ὠθέω, I thrust). As he also found that there was an outward flow of the liquid which increases in volume, or of some of its constituents, he attributed this to an opposite process, which he called "exosmose." For a time much confusion arose through a theory, introduced by Jolly, that a definite quantitative relation or "equivalence" exists between these two processes. It has been recognised for long that the passage outwards or inwards of dissolved constituents is simply due to diffusion; but the word osmosis remains to indicate the process by which water or any other solvent thrusts itself, or is attracted or driven, or otherwise penetrates, into a solution or another solvent. Round the originally convenient expression "osmosis," a dense haze of mystery and confusion has gathered, and has never been denser than in recent years.

---

[1] Dutrochet, *Mémoires pour servir à l'histoire anatomique et physiologique des végétaux et des animaux*, 1837.

Osmosis had from the time of Dutrochet onwards interested physiologists specially, as it gave promise of throwing much light on various physiological processes. Nägeli had clearly seen its relations to the varying "turgor" of vegetable cells[1]; and osmosis was freely invoked by animal physiologists in the hope of explaining various phenomena of absorption and secretion.

In 1861 Graham published the fact that vegetable parchment, though it allows water to pass, is completely impermeable to albuminous and various other organic and inorganic dissolved substances, which he thus distinguished as "colloids."[2] Their solutions give permanent and definite osmotic pressures, varying with the strength of the solutions and the nature of the dissolved substance; whereas the pressures given by membranes of bladder are not permanent, since even colloids leak through them. A membrane which stops a solute, but allows a solvent such as water to pass, is now always known as a semi-permeable membrane.

A further step forwards was taken when M.

---

[1] Nägeli and Cramer, *Pflanzenphysiologische Untersuchungen*, 1855, Heft I.

[2] Graham, *Phil. Trans.*, 1861, p. 183.

Traube discovered in 1867 that membranes can be formed which are impermeable, not merely to colloids, but also to many ordinary crystalloid substances.[1] One of the most effective of these membranes was obtained by dropping strong ferrocyanide solution into dilute copper sulphate solution. A membrane of copper ferrocyanide forms itself round the drops, which then swell up greatly, owing to the osmosis of water into them, while no ferrocyanide escapes.

Following up Traube's experiments, the vegetable physiologist Pfeffer devised a method of using a copper ferrocyanide membrane in such a way as to measure the osmotic pressures produced in solutions of substances to which the membrane is impermeable. Traube's copper ferrocyanide membranes could not be used for measuring osmotic pressure, as they are quite fragile; so Pfeffer introduced the plan of forming the membrane within the walls of a porous pot or tube capable of resisting heavy pressure (Fig. 4). By this means he found that when solutions of sugar or various other crystalloid substances were placed inside the porous pot permanent osmotic pressures up to several atmospheres could be obtained and measured with a suitable gauge.[2] The permanent pressure increased very exactly in proportion to the concentration of the solution with varying concentrations of cane sugar up to 6 per cent. This gave a pressure of about four atmospheres, which was about as much as the apparatus could stand without leaks. As the temperature rose the permanent pressure rose also. The concentrations by weight of the solutions of different substances which produced the same osmotic pressures varied, roughly speaking, in inverse proportion to their diffusibility. Thus, with a given concentration in water of a very indiffusible substance such as gelatin or gum arabic, the osmotic pressures were trifling in comparison with those produced by equal concentrations by weight of various salts or sugars; but in the case of most salts the full osmotic pressures could not be measured, as they leaked through the pot.

---

[1] M. Traube, *Arch. f. Anat. u. Physiol.*, 1867, p. 67: also *Botan. Zeitung*, 1875, p. 56.

[2] Pfeffer, *Osmotische Untersuchungen*, 1877.

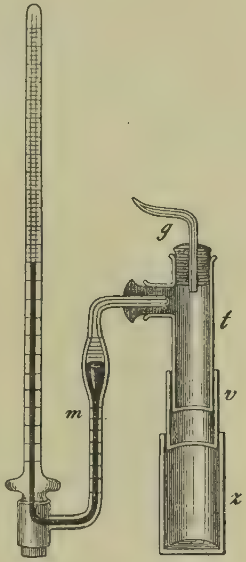

FIG. 4.—Pfeffer's Osmometer.

The semi-permeable membrane of precipitated copper ferrocyanide was deposited in the inner surface of the porous pot z. A glass cylinder v is fixed into the top of the pot with sealing-wax, and into v is sealed a glass T-tube with rubber corks through which connection is made, as shown, with an air manometer m and a glass tube g with its end drawn out. The latter is sealed after the whole apparatus has been filled with the solution to be tested. The osmotic pressure is indicated by the corrected manometer reading.

About the same time as Pfeffer was investigating osmotic pressure another botanist, de Vries,[1] was investigating osmosis quantitatively by a quite different method. When portions of living plant tissue are placed in sufficiently strong solutions, the layer of protoplasm which is normally applied to the inner surface of the external wall of a cell shrinks away from the wall, as shown in Fig. 5, owing to osmotic outflow of water from the space occupied by the protoplasm. De Vries called this phenomenon plasmolysis. As Nägeli had originally shown, the surface layer which bounds the protoplasm acts, like

---

[1] H. de Vries, *Archives Néerlandaises*, vi., p. 117, 1871: also "*Untersuchungen über die mechanische Ursachen der Zellstreckung*," 1877: also, in summary, *Zeitschr. f. physikalische Chemie*, II., p. 415, 1888.

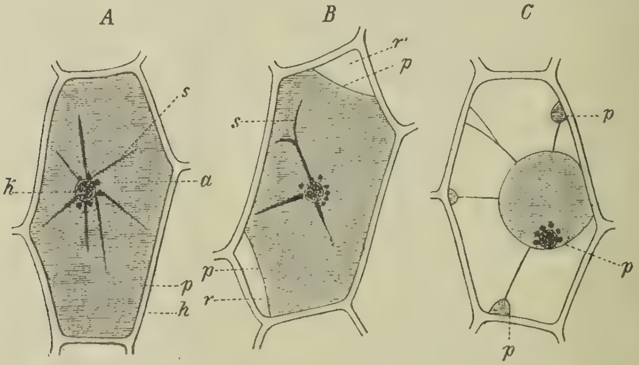

FIG. 5.—Cells from the epidermis of the central vein of a leaf of *Tradescantia discolor*. A, normal cell. B, plasmolysis in 0·22 mol. cane-sugar solution. C, extreme plasmolysis in 1·0 mol. potassium nitrate solution. *After* de Vries.

Dutrochet's membranes, so that osmotic interchange is set up, accompanied by swelling or contraction of the cells, or pressure in them, when the necessary conditions exist. This surface layer is evidently impermeable to many crystalloid solutes which go through a parchment membrane, and is in this respect similar to a copper ferrocyanide membrane. De Vries used plasmolysis as an index of the osmotic equivalence of different solutions of salts and various organic substances; and the very important results which he reached will be referred to below. He introduced the term "isotonic," now universally applied to solutions which produce equal osmotic effects.

The next great step was taken by van't Hoff.[1] Many of Pfeffer's pressure experiments were very carefully carried out, and, without making further experiments, van't Hoff, with the intuition of genius, saw that the pressures observed by Pfeffer with different concentrations of sugar, and at different temperatures, were, as nearly as could be expected, those which would be produced on the assumption that the sugar was present as a gas, for which the equation $pv = .082\,n\,\mathrm{T}$, or $\mathrm{PV} = \mathrm{RT}$, held good, so that osmotic pressure seemed to correspond to the bombardment pressure of gas at an ordinary pressure. In making this calculation, van't Hoff assumed that Pfeffer's concentrations could be taken as representing grammes per 100 c.c. of solution. It appears from Pfeffer's book that they were grammes per 100 grammes of solution, though van't Hoff himself inadvertently stated that they were grammes per 100 grammes of water. The differences between these different sorts of concentration were so small in Pfeffer's experiments that it was impossible to decide as to which, if any, of these varieties of concentration the observed variations of osmotic pressure corresponded to. Nevertheless, as we shall see, this question was of essential importance, and van't Hoff's assumption that in dilute solutions osmotic pressure is proportional to grammes per 100 c.c. of solution was incorrect.

---

[1] Van't Hoff, *Zeitschr. f. physikal. Chemie*, I., p. 481, 1887.

The experimental data quoted by van't Hoff from Pfeffer, and compared with what would theoretically be expected on van't Hoff's own interpretation, show how close the apparent correspondence was between the theory and actual observation. As regards variation of osmotic pressure with concentration of the solute, Pfeffer's experiments with sugar solutions at practically constant temperature gave the following results :—

Concentration per cent. by Weight.	Osmotic Pressure in mm. of Mercury.	Pressure Divided by Concentration.
1·0	535	535
2·0	1016	508
2·74	1518	554
4·0	2082	521
6·0	3072	512

These data correspond well with van't Hoff's theory.

On his interpretation, the osmotic pressure corresponded to the concentration of the solute per 100 c.c. of solution, in accordance with Boyle's law. Thus, if the osmotic pressure were an expression of the pressure of the solute molecules, Boyle's law in its original crude form would hold good in spite of the enormous concentration of water and sugar as compared with the concentration of a gas at anything like ordinary atmospheric pressure.

As regards the application of Charles's law to osmotic pressures, van't Hoff showed that Pfeffer's results gave perfectly clear indications. The osmotic pressure varied as nearly as could be expected with the absolute temperature, as shown by the following results for 1 per cent. sugar solutions:—

Temperature °C.	Osmotic Pressure in mm. Hg.	
	Observed.	Calculated from Charles's Law.
32°	544	...
14·15°	510	512
36°	567	...
15·5°	520·7	529

As regards the applicability of Avogadro's law to osmotic pressure, the agreement between observation and calculation was still more striking, and left not the slightest doubt as to the connection between osmotic pressure and the gas laws. Van't Hoff gives the following table comparing Pfeffer's results for a 1 per cent. solution of sugar and the pressure which would be expected if osmotic pressure were due to bombardment pressure of solute molecules, and if not

Temperature °C.	Osmotic Pressure in Atmospheres.	Gas Pressures at the same Molar Concentration if Boyle's, Charles's, and Avogadro's Law Applied.
6·8°	·664	·665
13·7°	·691	·681
14·2°	·671	·682
15·5°	·684	·686
22°	·721	·701
32°	·716	·725
36°	·746	·735
Mean . .	·699	·696

only Boyle's and Charles's, but also Avogadro's law held good in its original crude form as applied to gases at ordinary pressures.

Van't Hoff then took the concentrations which de Vries had found by the plasmolysis method to give isotonic solutions in the case of a number of organic substances. The result of the examination was to show that with cane sugar, invert sugar, malic, tartaric and citric acids, as well as malate and sulphate of magnesia, the solutions possessing the same number of moles per litre were isotonic. Avogadro's law could thus be extended to the osmotic pressures given by many different solutions, though in the case of various inorganic salts there were anomalies which will be referred to later. It must be remembered, however, that what de Vries had measured was not osmotic pressure, but isotonicity.

In view of the close correspondence between osmotic pressure and what he calculated would be the gas-pressure of the solute, van't Hoff concluded that since gas-pressure is just molecular bombardment pressure, osmotic pressure must be due to the bombardment pressure of the solute molecules. As to the connection between osmotic pressure and the gas laws, van't Hoff's argument left not a shadow of doubt. The theory generally adopted up to that time that osmosis is due to an attraction of the solute molecules for water was therefore abandoned. But it is very strange that van't Hoff was contented with the theory that the bombardment pressure of the solute molecules causes osmotic pressure.

This theory is inherently unintelligible. The bombardment pressure of the sugar in solution is certainly real; but its action can only be (and actually is) to keep water out of the solution, and not to drive water in. The pressure is the wrong way for van't Hoff's explanation, as indeed has often been pointed out. But the theory would not have mattered so much if the facts which it was an attempt to embody had been correct. Later experiments have shown that it is not true, whether for dilute or for strong solutions, that the osmotic pressure, as measured, varies with the molar concentration per unit of volume, though this is approximately the case with very dilute solutions.

The experimental difficulties of work on osmotic pressure are considerable on account of the very high pressures concerned, and the consequent trouble in keeping the apparatus tight. Hence Pfeffer was only able to work on comparatively dilute solutions, giving pressures up to only six atmospheres; and owing to imperfect rigidity of the apparatus the higher measurements of pressure seem to have been too low.

As is well known, these difficulties have been to a great extent overcome by the skill and patience of the late Professor Morse and of the Earl of Berkeley, and their associates. Morse's results showed that at equal temperatures the osmotic pressure of an aqueous solution varied, roughly speaking, in proportion, not to moles of solute per litre, but to moles per 1000 grammes of water present. In hot solutions of cane sugar the correspondence of the experimental results with this ratio was extremely close. Morse therefore called a solution "weight-normal" when it contained, not one mole per litre, but one mole per 1000 grammes of solvent. He concluded that the *n* in van't Hoff's equation for osmotic pressure ($P_0 = nRT$, when $n =$ number of moles of solute per litre) means, not moles of solute per litre of solution, but per litre of water at 4° C., or per 1000 grammes of water. This interpretation has been generally accepted, though it is certainly a most extraordinary one; and how it could possibly be linked up with the dynamical theory of gases has been left totally obscure.

At temperatures below 80° in the Morse experiments the osmotic pressure observed did not coincide with the "weight-normal" concentration very satisfactorily for sugar solutions; and this lack of agreement was brought out much more clearly in the experiments of Berkeley and Hartley, who still further improved the method, and succeeded in measuring osmotic pressures up to 135 atmospheres.

These experiments will be discussed below; but meanwhile it is necessary to return to the comparison between osmotic pressure and the calculated bombardment pressure of the sugar or other solute in an aqueous solution.

When we compare a liquid with a gas, it is evident that the molar concentration in the liquid is enormously greater than in the same substance in the ordinary gaseous state at the same temperature. For instance, one litre of steam at 100° C. and ordinary atmospheric pressure would, if it were a perfect gas, contain only

.0327 moles per litre, whereas one litre of water at the same temperature contains 53.25 moles, or 1633 times as many. For this reason alone, therefore, the molecular bombardment pressure within the water must, in accordance with the dynamical theory, be at least 1633 atmospheres. But since $P(1 - Nb) = N.082T$, and since $1 - Nb$ in the case of water or other liquids is far less than one, the value of P must be far greater than 1633 atmospheres. Van der Waals calculated it as 10,500 atmospheres for water. If, as will be shown later to be justifiable, we assume that the value of P is the same in pure water and in aqueous solutions of sugar, the bombardment pressure of the sugar when 1 mole of sugar per litre is present so that $n = 1$, and the total moles in a litre of pure water are N, and in a litre of solution N_1, will be $\frac{N - N_1 + n}{N} P$, as was shown in Chapter I. Even if we took the partial pressure as only $\frac{n}{N} P$, when $n = 1$, it would amount at 0° C. to $\frac{1}{55.5} \times 10,500 = 196$ atmospheres, which is about nine times the osmotic pressure calculated on van't Hoff's theory. Since, however, one mole of cane sugar displaces many moles of water in the solution, the value of N_1 in a molar solution of cane sugar is considerably less than that of N, and this still further increases the disparity. The true bombardment pressure of the sugar is thus enormously higher than corresponds to the n in van't Hoff's equation.

Van't Hoff's theory that osmotic pressure varies, at whatever dilution, with the molar concentration and bombardment pressure of the solute molecules has thus two defects, either of which wrecks the theory. (1) The osmotic pressure does not, as a matter of fact, vary in proportion to the concentration of the solute molecules, but in proportion, roughly speaking, to the ratio of solute molecules to mass of water present. (2) The osmotic pressure is a mere fraction of the bombardment pressure of the solute and does not vary in proportion to this bombardment pressure.

Apart from these quantitative defects the theory is quite unintelligible, in spite of various totally unsatisfactory attempts which have been made since van't Hoff's time to render it intelligible. There is no conceivable reason why the bombardment pressure of solute molecules within a solution should drive water *into* the solution. To turn to physiology, there is also no intelligible reason why a very minute deficiency in the combined bombardment pressures of solute molecules in the blood should act, as it appears on van't Hoff's theory to do, as a powerful stimulus to the excretion of water by the kidneys.

It is not easy to understand how van't Hoff was so strangely misled in his interpretation of osmotic pressure; but probably his error was largely due to trusting too much in mathematical formulæ without considering their physical meaning and the limits of their applicability. A mere difference in sign, such as distinguishes a positive from a negative change in pressure, might not cause so much trouble from the mathematical as from the physical side. The equation $PV = RT$ might also appear from the abstract mathematical side as a weapon capable of being wielded in all directions, apart from the experimental observations which it embodies.

Comparing the pressure of a gas with the osmotic pressure of a solution, van't Hoff remarks that, "We are concerned in the former case with the impacts of gas molecules on the wall of the containing vessel, in the latter case with the impacts on the semi-permeable membrane of the solute molecules, since the molecules of the solvent, present on both sides of the membrane, and passing through it, do not come into account." Taking this suggested account as it stands, it is evident that there would be a positive pressure on the membrane when a solution was placed in an osmometer. This would tend to drive water out, just as when there is mechanically produced pressure in the osmometer. But actually the water comes in. Moreover, we cannot disregard the bombardment pressure of the solvent, since this is less in the solution, where the solvent is diluted by the solute, than in the pure solvent. It may safely be said that since the times of the phlogiston and caloric theories no more confused hypothesis has passed into general currency than van't Hoff's theory of osmotic pressure. It is none the less true that, like the phlogiston and caloric theories, van't Hoff's theory has been extraordinarily fruitful, and has for many purposes been near enough to truth.

The explanation to which the quantitative relations connected with osmotic pressure point very clearly will now be evident enough in view of the discussion in Chapter III. (p. 120). We saw there that when a gas is separated by a septum exclusively permeable to this gas from a mixture of the same gas with another gas at the same pressure, P, measured in atmospheres, the osmotic pressure, P_0, equals $\frac{n}{N_1 - n} P(1 - Nb)$, which equals $\frac{nN}{N_1 - n} .082\,T$, where N, N_1, and n have the significances specified at page 22. This osmotic pressure is the additional external pressure which must be applied to the mixture in order to equalise on the two sides the diffusion pressures of the gas which passes the septum. We also saw that at the very high bombardment pressures of the liquid state the pressure is the same whether it is produced by substance passing into the mixture (as in Pfeffer's and Morse's experiments) or by external pressure directly applied (as in the Earl of Berkeley's experiments). If, now, we assume that as regards diffusion the two liquids concerned in liquid osmosis behave in the same way as very highly compressed gases at the *same* initial bombardment pressure, we can test this assumption by the actual quantitative relations between osmotic pressure and concentration. Comparison of the theory with the facts is the more easy since, as was shown at page 122, $\frac{nN}{N_1 - n}$ represents moles of solute per 1000 grammes of solvent when, like water at or near 4°, the solvent has a specific gravity of 1. For the present we may postpone till Chapter V. the further discussion of the implications involved in the assumption, and simply examine whether the observed quantitative relations in question correspond to the assumption. The results of Berkeley and E. G. J. Hartley[1] will be taken first, since their measurements extended to very high pressures, and, as will be shown below, were more accurate than most of those of Morse. Pfeffer's results were probably too low, on an average, at his higher pressures, because of the connections in his apparatus being insufficiently rigid.

To overcome the difficulties connected with measuring very high pressures without letting water leak in, Berkeley and Hartley placed the water inside the osmometer, and used a capillary gauge as an index of leakage of water inwards or outwards. By this means, and by using an exactly calibrated Bourdon gauge, great accuracy of measurement at high pressures could be assured.

The appended table (page 146) shows the results obtained by Berkeley and Hartley for cane-sugar solutions in water at 0° C., together with the concentrations in grammes and moles per litre and per 1000 grammes of water. The osmotic pressures as calculated by different equations are also given. It was suggested by Callendar[2] that the reason why the osmotic pressure of a cane-sugar solution does not vary in Morse's "weight-normal" proportion is that, although no cane-sugar hydrates have hitherto been isolated, the sugar is nevertheless present in solution

---

[1] The Earl of Berkeley and E. G. J. Hartley, *Proc. Roy. Soc.*, A, 92, p. 483, 1916.

[2] Callendar, *Proc. Roy. Soc.*, A, 80, p. 466, 1908.

Osmotic Pressures of Cane-Sugar Solutions at o° C.

Weight Concentration.		Volume Concentration.		Osmotic Pressure in Atmospheres.			
Grammes $C_{12}H_{22}O_{11}$ per 100 g. H_2O.	Grammes $C_{12}H_{22}O_{11}+5H_2O$ per 100 g. free H_2O.	Grammes $C_{12}H_{22}O_{11}$ per 100 c.c.	Moles Sugar per litre.	Actually observed by Berkeley and Hartley.	Calculated from the equation $P_0=\frac{n}{N_1-n}\times\cdot082\,NT$ assuming that the Sugar is present as Pentahydrate.	Calculated from the same equation, assuming that the Sugar is not Hydrated. (Morse's Theory.)	Calculated from the equation $P_0=\cdot082\,nT$. (Van't Hoff's Theory.)
3·393	4·40	3·32	0·097	2·23	2·23	2·22	2·17
10·18	13·57	9·59	0·281	6·85	6·87	6·66	6·29
20·525	28·21	18·26	0·533	14·21	14·29	13·44	11·95
30·61	43·04	25·81	0·754	21·87	21·82	20·04	16·90
33·945	48·64	28·13	0·822	24·55	24·64	22·22	18·41
81·26	136·2	54·24	1·585	67·74	68·99	53·19	35·48
111·73	214·4	66·14	1·933	100·13	108·6	73·14	42·20
141·11	310·7	75·33	2·201	134·84	157·5	92·37	49·31

as a pentahydrate. The grammes and moles of sugar per 1000 grammes of free solvent are therefore also given on this assumption.

On examining the table it will be seen that even when the concentration was only 3.32 grammes per 100 c.c., or .097 moles per litre, the osmotic pressure as calculated on van't Hoff's assumption that $P_0 = n\ .082\ T$ was 3.1 per cent. too low, and that with increasing concentration it became proportionally lower and lower than the actual value, and was 63 per cent. too low at the highest concentration investigated.

The osmotic pressure calculated on Morse's theory that the osmotic pressure is .082 T multiplied by the number of moles present per 1000 grammes of water, the sugar being taken as unhydrated, came much nearer the actual values, but was nevertheless already nearly 3 per cent. too low with a solution containing 9.59 grammes per 100 c.c., and was 31 per cent. too low with the highest concentration. When, on the other hand, P_0 is calculated from the theoretical equation given above, but assuming the sugar to be present as pentahydrate, it will be seen from the table that the agreement between calculated and observed result is extraordinarily close until a pressure approaching 70 atmospheres is reached. At the highest pressure, however, the equation gives a result about 17 per cent. too high.

The explanation of the discrepancy at very high concentrations is possibly the fact that at such great concentrations the molecules of the sugar are becoming, to a very appreciable extent, unhydrated. The fact that sugar crystallises out in the unhydrated form from a super-saturated solution supports this supposition and points to the conclusion that in a saturated solution there might be no hydration at all. There is, however, another and more probable explanation of the discrepancy. As will be pointed out in Chapter VII. (page 276), it is impossible that in the case of a solution of a solute with a comparatively high melting-point, like sugar, in a solvent of low melting-point, like water, the bombardment pressure in a very strong solution can, with very high concentrations, remain as low in the solution as in the pure solvent. This would account at once for the discrepancy and its direction.

The exact agreement between the observed osmotic pressures and those calculated on the pentahydrate theory seems to leave no doubt that at 0° C. the sugar is actually present, up to high concentrations, wholly as pentahydrate. The agreement is at the same time a tribute to the accuracy of Berkeley and Hartley's work, and to its very great scientific value. Without this work a correct theory of osmotic pressure could not have been verified satisfactorily by the results of experiments.

The results given in the preceding table may be stated in a different form. If the osmotic pressure equals $\frac{nN}{N_1 - n}$ .082 T, it is evident that this amount divided by $\frac{nN}{N_1 - n}$ T must equal .082. The quotient may be called the molar osmotic pressure, but must be distinguished from the molar or molecular osmotic pressure as calculated on van't Hoff's theory that the osmotic pressure varies as the molar concentration per litre of the solution. The following table gives the molar osmotic pressure as calculated from Berkeley and Hartley's experiments on cane-sugar solutions at 0° C., together with the molar osmotic pressures as calculated on the theories of van't Hoff and Morse. This table speaks for itself.

Osmotic Pressures of Cane Sugar at 0° C.

Grammes Anhydrous Sucrose per 100 g. of Water.	Moles Sucrose Pentahydrate per 1000 g. Free Water.	Observed Osmotic Pressure (Berkeley and Hartley) in Atmospheres.	Molar Osmotic Pressure.		
			Calculated from data in Column 3.	Calculated from Columns 1 and 3 on Morse's Principle.	Calculated directly from Molecular Concentration (see preceding Table) on Van't Hoff's Principle.
3·393	0·0996	2·23	·082	·082	·084
10·18	0·307	6·85	·082	·084	·089
20·525	0·638	14·21	·082	·087	·098
30·61	0·974	21·87	·082	·089	·106
33·945	1·100	24·55	·082	·091	·109
81·26	3·080	67·74	·081	·105	·157
111·73	4·851	100·13	·076	·112	·190
141·11	7·030	134·84	·068	·120	·230

The next table embodies the results of Berkeley and Hartley for dextrose.[1] Unfortunately they only give the concentrations in grammes per litre. The concentrations in moles per litre are calculated from the specific gravity table quoted from Morse in Landolt and Börnstein's *Physikalisch-chemische Tabellen.* It seems evident from the table that the dextrose is present in solution as a di-hydrate; and, just as in the case of cane sugar, the internal bombardment pressure of the solution is apparently increased to some extent at the highest concentration investigated.

---

[1] The results for dextrose and galactose are from an earlier paper (*Phil. Trans.*, A, 206, p. 481) and are probably not so correct as those first quoted.

Osmotic Pressures of Dextrose Solutions at 0° C.

Grammes Dextrose per litre of Solution.	Moles Dextrose per litre of Solution.	Moles Anhydrous Dextrose per 100 g. of Water.	Moles Dextrose Di-hydrate per 1000 g. Free Water.	Observed Osmotic Pressure in Atmospheres.	Calculated Molar Osmotic Pressure.		
					For Dextrose Di-hydrate (Columns 4 and 5).	For Anhydrous Dextrose (Columns 3 and 5). Morse's Principle.	For Molar Concentration (Columns 2 and 5) Van't Hoff's Principle.
99·8	0·554	0·591	0·604	13·21	·080	·082	·087
199·5	1·107	1·265	1·327	29·17	·081	·084	·096
319·2	1·772	2·215	2·407	53·19	·081	·088	·110
448·6	2·491	3·454	3·948	87·87	·082	·090	·129
548·6	3·046	4·669	5·612	121·18	·079	·096	·146

Berkeley and Hartley determined also the osmotic pressures of several solutions of galactose $(C_6H_{12}O_6)$, and the next table contains their results. As before,

Osmotic Pressures of Galactose Solutions at 0° C.

Grammes Galactose per litre of Solution.	Moles Galactose per litre of Solution.	Moles Anhydrous Galactose per 1000 g. of Water.	Moles Galactose Di-hydrate per 1000 g. Free Water.	Observed Osmotic Pressure Atmospheres.	Calculated Molar Osmotic Pressure.		
					For Galactose Di-hydrate (Column 4).	For Anhydrous Galactose (Columns 3 and 5). Morse's Principle.	For Molar Concentration (Columns 2 and 5) Van't Hoff's Principle.
250	1·388	1·61	1·70	35·5*	·076	·081	0·94
380	2·110	2·59	2·85	62·8	·081	·089	1·09
500	2·776	3·70	4·27	95·8	·082	·095	1·26

---

* There was only one observation at the lower concentration, and the recorded pressure seems to have been too low.

I have calculated the "weight normal" concentrations, but from rather imperfect specific gravity data, since the authors did not give the specific gravities or "weight normal" concentrations of their solutions. From column 6 it appears that the sugar is present as a di-hydrate.

Morse summarises the results reached by him and his co-workers for cane sugar in the following table.[1] In this table the ratios are given of the observed osmotic pressures to the "gas pressures" of the sugar. Morse clung to van't Hoff's theory that osmotic pressure represents gas pressure of the solute, but was forced to assume (in contradistinction to van't Hoff) that the volume into which the sugar solution is compressed is not that of the solution, but that which would be occupied by 1000 grammes of the solvent at 4° C. (the temperature of maximum density of water). This comes to the same thing as assuming that the osmotic pressure varies as moles of solute per 1000 grammes of solvent. The calculated osmotic pressure, or gas pressure, is thus not based simply on the equation $P_0 = \frac{nN}{N_1 - n} .082\,T$, but on the equation $P_0 = \frac{nN}{(N_1 - n)S} .082\,T$, where S is the specific gravity of the pure solvent, not at the temperature of the observation, but at 4°.

---

[1] The experiments of Morse and his colleagues appeared in the *American Chemical Journal* for 1901, 1902, 1903, 1904, 1905, 1907, 1908, 1909, 1911 and 1912.

Ratios of Osmotic Pressure to Gas Pressure in the Case of Cane-Sugar Solutions.

Weight Normal Concentration.	Ratio of Osmotic Pressure to Gas Pressure at											
	0°	5°	10°	15°	20°	25°	30°	40°	50°	60°	70°	80°
0·1	(1·106)	1·082	1·082	1·082	1·084	1·084	1·000	1·003	1·000	1·000	...	...
0·2	1·061	1·063	1·060	1·061	1·062	1·059	1·020	1·011	1·002	1·001	...	...
0·3	1·061	1·058	1·059	1·061	1·060	1·060	1·031	1·024	1·009	0·999	...	...
0·4	1·060	1·059	1·060	1·059	1·060	1·059	1·040	1·038	1·017	1·000	...	...
0·5	1·069	1·067	1·066	1·068	1·067	1·065	1·050	1·046	1·025	1·006	1·000	...
0·6	1·077	1·074	1·073	1·073	1·073	1·071	1·060	1·054	1·032	1·015	1·002	...
0·7	1·083	1·084	1·083	1·083	1·084	1·083	1·069	1·059	1·041	1·020	0·999	...
0·8	1·093	1·093	1·092	1·093	1·093	1·093	1·081	1·067	1·049	1·027	1·008	1·001
0·9	1·104	1·102	1·102	1·102	1·103	1·102	1·089	1·076	1·059	1·033	1·015	1·000
1·0	1·115	1·115	1·113	1·115	1·115	1·113	1·101	1·085	1·071	1·044	1·023	1·000

The vertical line in the table separates the results of earlier experiments (to the left) from later experiments carried out at higher temperatures and with improved technique. On examining the figures to the left of the line it will be seen what led Morse to calculate his "gas pressures" in the manner just described. In the first place, if he had calculated the "gas pressures" on van't Hoff's assumption that the pressure varies as the sugar concentration per litre, there would have been extremely bad agreement between theory and observation, as has already been shown in connection with the results of Berkeley and Hartley. The agreement was far better on the basis of Morse's theory. On running the eye horizontally along the columns it will be seen that the apparent agreement with Charles's law is extremely close. On running the eye vertically up and down the columns it will also be seen that there is a fair apparent agreement with Boyle's law. On the other hand, the osmotic pressures are on an average about 8 per cent. higher than they would be if Avogadro's law held good. Morse suggests that this is due in some way[1] to hydration of the sugar; and this suggestion seemed to be completely confirmed when he found that at 80° the discrepancy had disappeared, since hydration would naturally diminish with rise of temperature.

When, however, we examine the whole table, several anomalies become evident. In the first place, there is a sudden drop between 25° and 30° in the observed osmotic pressures at the lower concentrations. If this were due, as Morse suggested, to a sudden diminution in hydration, owing to the rise from 25° to 30° having a critical effect, we should expect the drop to be equally marked at the higher concentrations. This, however, was not the case. Moreover, on looking up and down, or along, the columns to the right of the vertical line we see that the results agree with neither Boyle's nor Charles's nor Avogadro's laws. The only assumption that would explain the discrepancies (assuming the experimental results to be correct) is the extremely improbable one that in the dilute solutions there was no hydration at all, even at 30°, while in the strong solutions there was considerable hydration, even at 70°.

The unsatisfactory agreement between Morse's results and theory directs attention to possible sources of fallacy in his measurements, as well as in his calculations. To take the measurements first, the line which sharply separates the older from the newer experiments, and the equally sharp difference in results on either side of this line, suggest that a change in experimental method might account for the difference. The new experiments were made with thoroughly matured and seasoned cells, the membrane being in some cases of copper ferrocyanide, and in other cases of nickel ferrocyanide. In the earlier experiments the precaution was taken of placing what were believed to be isotonic (iso-ionic) dilute solutions of ferrocyanide of potassium and copper sulphate inside and outside the cells, so that if any leakage developed in the membrane the leak would be repaired by deposition of fresh copper ferrocyanide. No experimental proof was obtained, however, that these solutions were actually isotonic under the conditions of the experiment, or that their presence did not affect the results obtained with the sugar. It is probable, as will be shown below, that an error affecting very appreciably the results with the more dilute solutions of sugar was caused by the presence of these salts. In the later experiments at higher temperatures no salts were used, as the membranes were in all cases well matured.

---

[1] On the theory of van't Hoff, to which Morse clung, hydration ought to have no effect on the osmotic pressure.

In any case, there can be no doubt that an experimental error of some kind had affected the older results for the lower concentrations of sugar. Berkeley and Hartley, who worked with no salts present, compared their results with those of Morse at equal concentrations of sugar, with the following results calculated in the same way as Morse calculated.

Ratio of Observed Osmotic Pressure to Calculated Gas Pressure.

Cane-Sugar Solutions at 0° C.

"Weight-normal" Concentration, assuming no Hydration.	Ratio from Berkeley and Hartley's Results.	Ratio from Morse's Results.
·0992	1·008 [1·012]	1·098 [1·082]
·2978	1·031	1·057
·6001	1·062	1·075
·895	1·096	1·099
·993	1·109	1·110

It will be seen that at the lowest concentration Morse's results were about 7 per cent. higher, but that as the concentration increased the difference diminished till it became inappreciable at the highest concentration. The figures enclosed in brackets are probably more nearly correct than those actually given for the same concentration. If there were any doubt as to Berkeley and Hartley's results being the correct ones, this is removed by the fact that, as will be shown in Chapter IV. (page 178), they agree exactly with entirely independent figures for depression of freezing-point.

The corrected figures fail, of course, to show the agreement which Morse thought he had found with the gas laws. This is explained, when the results are correctly interpreted, by the sugar being present as pentahydrate; but other minor anomalies in Morse's results still remain. If the results are calculated out on the assumption that the theoretical osmotic pressure, or "gas pressure," is $\frac{nN}{N_1 - n}$ .082 T it will be found that, roughly speaking, where Morse gives the ratio of observed to theoretical pressure as 1.000, the ratio given by the expression just quoted is 1.005. I at first thought that Morse must have taken the figure .0816 instead of .0820 in making his calculations of the theoretical pressure; but this was not the case. On looking, however, at his method of measuring osmotic pressure I found that he had measured the pressure with a pressure-gauge filled with nitrogen, but had made no allowance for the changes in the value of pv in the case of nitrogen with increase of pressure. These changes were determined exactly by Amagat in his famous experiments at the Verpilleux Colliery and elsewhere in 1879. Since at ordinary temperatures the value of pv for nitrogen diminishes appreciably and steadily up to the pressures measured by Morse, it follows that his estimate of the osmotic pressures was appreciably too high at the higher pressures. Thus at a weight-normal concentration of 1 there was in reality a slight difference (of about 0.6 per cent.) between his measurements and those of Berkeley and Hartley at 0°. This is what one would expect on the theory that Morse's measurements were vitiated owing to the presence of the salts referred to above. In correcting his manometer readings Morse simply used $.00367t$ as the correction factor. This, however, is applicable only to air at constant pressure. With both temperature and pressure varying, the proper correction is much more troublesome, and in the absence of such correction Morse's results do not possess the degree of extreme accuracy which his tables at first suggest, although it is to the experimental skill and enthusiastic persistency of Morse and his associates that we owe much of the present knowledge as to osmotic pressure.

A consideration not yet taken into account makes Morse's results at the higher temperatures much more intelligible. As his table stands, it appears that, to take for instance the results at 60°, there is no hydration of the sugar in the more dilute solutions, but considerable hydration in the stronger ones. This is quite unintelligible. Now Morse calculated his osmotic pressures or "gas pressures," not from the correct equation $P_0 = \frac{nN}{(N_1 - n)S} .082\ T$, where S is the specific gravity of water at the existing temperature (see p. 112), but from the same equation where S is the specific gravity at 4° C. At the lower temperatures the value of S is almost exactly 1 in any case; but at 60°, for instance, it is .983 at the existing temperature. In reality, therefore, his ratio for a 0.1 weight-normal sugar solution was 1.7 per cent. too high, though this excess is reduced if the gauge is properly corrected. The full hydration effect seems, therefore, to have been present. It appears probable that even up to 80° there was appreciable hydration in the more dilute solutions, though only very slight hydration in the more concentrated ones. This is just what might be expected.

Morse's experiments are of special value because they were made at so considerable differences of temperature. When due allowance is made for errors of experiment or of calculation these experiments show in the clearest way that osmotic pressure varies, within considerable limits of temperature, almost exactly as the absolute temperature, multiplied by the molar concentration of the pure solvent, in accordance with the theory that P, the internal bombardment pressure, was the same in the pure solvent and the solution, not merely at one temperature, but also at the different temperatures investigated: in other words, that the osmotic pressures increased in the same way as they would in a gas at very high pressure if the pressure continued to be the same on the two sides of the semi-permeable membrane. The explanation of the equal internal pressures in pure solvents and solutions in them must be postponed to Chapter VII.

It is unfortunate that the existing data for osmotic pressures with solutions of non-electrolytes are so scanty. There are, it is true, a few data for colloids and for dyes in enormous molecular dilution; but the pressures produced were too small to be of any value in deciding the exact form of the equation for osmotic pressure. Moreover, the results were sometimes also very seriously vitiated through the presence of electrolytes inside and outside of the osmotic cell.

Nevertheless, the existing results with sugar solutions show quite definitely that, as regards osmotic pressure, sugar solutions and water behave exactly like gases at the same pressure, provided that this is a sufficiently high one.

In accordance with his theory of osmotic pressure van't Hoff spoke of osmotic pressure as being something which exists in solutions even when no external pressure is applied; and the scientific world has become just as much accustomed to this mode of expression as it was in the eighteenth century to expressing chemical changes in terms of the phlogiston theory. But according to the theory which has been developed in the preceding pages osmotic pressure is an externally applied pressure, and is the pressure which must be applied to a solution in order to equalise the diffusion of solvent from the solution into pure solvent with the diffusion of solvent from pure solvent into the solution. Hence, what is really meant by the "osmotic pressure" existing, apart from externally applied pressure, in a solution must be something different from actual osmotic pressure. It is evidently a condition in virtue of which solvent tends to pass into the solution; and in the light of the new theory we can see quite clearly what this condition is.

The solvent tends to pass in because the diffusion pressure of solvent from the solution into pure solvent or a weaker solution is less than the diffusion pressure of solvent from pure solvent or from the weaker solution into the solution. The "thrusting," which the word osmosis signifies etymologically, is just the thrusting of a greater diffusion pressure of solvent. The solution has a less thrusting power than the pure solvent or the weaker solution. It is the depression or deficiency of thrusting or diffusion pressure of solvent in the solution that is really meant when we ordinarily speak of the "osmotic pressure" existing in a solution. We can properly speak of the osmotic depression or deficiency of a solution; but to speak of an osmotic "pressure" existing in it when no external pressure is applied can only produce confusion; and endless confusion has, in fact, been produced in this way. When we mean osmotic depression we ought to use the expression "osmotic depression" and not "osmotic pressure." The osmotic depression in a solution, as in a gas, will be designated as P_d.

In Chapter III. (page 114) it was shown for gases at the same pressure, but separated by a semi-permeable membrane, that just as $P_o = \frac{nN}{N_1 - n} .082\ T$, so $P_d = \frac{nN}{N_1} .082\ T$, where n, N, and N_1 represent respectively the molar concentrations of gas which cannot penetrate the septum, of the pure gas which can penetrate, and of the mixture. Applying the same reasoning to liquids, and taking n, N, and N_1 as applying, respectively, to solute, pure solvent and solution, we obtain, as already shown, the actual osmotic pressure, and by similar reasoning we also obtain the osmotic depression P_d. It is therefore clear that P_d varies as the ratio of moles of solute which cannot penetrate the membrane to total moles in a given volume or weight of solution. It is also clear that when the ratio $\frac{n}{N_1}$ is equal in two solutions in the same solvent they will be isotonic, but not if this ratio is unequal. It is thus the value of the ratio, and not (as van't Hoff thought, and as is generally assumed at present) the value of n, that determines isotonicity. A solution of dextrose, for instance, would not be isotonic with a solution of sucrose of equal molar concentration. The concentration of free water molecules would be greater in the former than in the latter, so that the value of $\frac{n}{N_1}$ would be less in the former than in the latter, and water would pass from the dextrose into the sucrose solution. Similarly, blood or blood-plasma would not be isotonic with a salt solution of the same molecular (or ionic) concentration, but with a salt solution of considerably higher molecular concentration. This at once renders intelligible, for instance, the fact that when water containing only the same proportions of salts as the blood-plasma is drunk there is a great excretion of water by the kidneys. To prevent this excretion the concentration of salts in the water drunk must be higher.

A thorough clearing up of the existing confusion between osmotic pressure and osmotic depression is a matter of as great importance to the free development of physical chemistry as, in general chemistry, was the clearing up of the confusion associated with the phlogiston theory. Osmotic depression varies, as will be shown in Chapter V., parallel with depression of vapour pressure, while this is not the case for osmotic pressure. With the disappearance of the confusion between osmotic depression and osmotic pressure endless confusions and meaningless mathematical complications disappear from physical chemistry and also from physiology. We can also see that isotonicity in two solutions with the same solvent signifies simply equality, in the two solutions, of the mutual diffusion pressures of the *solvent*.

We must now turn to the osmotic pressures given by solutions of salts and other electrolytes. As already mentioned, Pfeffer was unable to measure the full osmotic pressures given by solutions of inorganic salts, as they leaked through his copper ferrocyanide membrane. His successors have fared no better in this respect. The comparative osmotic reactions given by a number of different solutions, including salt solutions, had already, however, been measured by de Vries, by means of his plasmolysis method, already described above. Using this method, he determined what concentrations of different solutes were isotonic with one another. On comparing these isotonic concentrations with the concentrations which would be expected if equimolar solutions gave the same osmotic pressures, van't Hoff found that solutions of salts gave osmotic reactions which were higher, and often much higher, than the expected reactions. Thus de Vries found in four series of experiments that solutions of the following molar concentrations were isotonic :—

Series.	Potassium Nitrate (KNO_3).	Cane Sugar. ($C_{12}H_{22}O_{11}$).	Potassium Sulphate (K_2SO_4).
1	0·12	...	0·09
2	0·13	0·2	0·1
3	0·195	0·3	0·15
4	0·26	0·4	...

When the molar concentrations of potassium nitrate were in each series taken as unity, the molar concentrations of the sugar and potassium sulphate were as shown in the next table :—

Series.	KNO_3.	$C_{12}H_{22}O_{11}$.	K_2SO_4.
1	1	...	·73
2	1	1·54	·77
3	1	1·54	·77
4	1	1·54	...

Van't Hoff did not himself discover the reason why the salts produced such unexpectedly high osmotic reactions, but he mentioned that an explanation had been suggested in a letter to him from Svant Arrhenius, a young Swedish investigator. A paper by Arrhenius followed in the next number of the same famous first volume of the *Zeitschrift für physikalische Chemie.*

When a gas at near atmospheric pressure gives a greater pressure or volume than it apparently ought to do in accordance with the gas laws, this has been found to be due to molecular dissociation. For instance, when bromine gives a higher value for PV than corresponds with the gas laws as applied to supposed molecules of Br_2, this is because the molecules of Br_2 are partially dissociated into monatomic molecules of Br. Arrhenius brought forward clear evidence that when salts give a higher osmotic pressure than corresponds to their calculated molecular concentration, this can be accounted for by dissociation of the solute molecules into the fragments which Faraday first recognised in solutions of electrolytes and which he named "ions," since they are the moving or going carriers of electric charges in the process of electrolysis. The conductivity of a solution must depend upon the degree of ionization of the solute molecules, and Arrhenius had already, in an earlier paper, shown how the percentage ionization of the solute molecules can be calculated from measurements of the conductivity of the solution and from its concentration.

He now brought forward full evidence that the percentage ionization as determined by conductivity measurements corresponds closely to the percentage increase in the osmotic reaction of electrolytes above the value calculated on the theory that the molecules of the electrolyte are not dissociated. As will be fully discussed in Chapter V., Raoult had discovered that the depression in freezing-point and raising of boiling-point of solutions of non-electrolytes is proportional to the "concentrations" of the solutions, and van't Hoff concluded by thermodynamical reasoning that depression of freezing-point, depression of vapour pressure, and osmotic pressure must run parallel to one another. The depressions of freezing-point in aqueous solutions of electrolytes were apparently abnormally great, just as were the osmotic pressures. Arrhenius compared his own determinations of percentage ionization with Raoult's determination of the abnormally increased depression of freezing-point in very dilute solutions of electrolytes (one gramme per 1000 grammes of water), and found that there was a striking correspondence between the two sets of figures. The accompanying table gives a few of the data which Arrhenius cited.

	Molar Depression of Freezing-point.	
	Calculated from Actual Depression.	Calculated from Percentage Ionisation
Cane sugar, $C_{12}H_{22}O_{11}$	1·00	1·00
Barium hydrate, $Ba(OH)_2$	2·69	2·67
Calcium hydrate, $Ca(OH)_2$	2·59	2·59
Sodium hydrate, $NaOH$	1·96	1·88
Potassium hydrate, KOH	1·91	1·93
Ammonia, NH_3	1·03	1·01
Hydrochloric acid, HCl	1·83	1·81
Hydriodic acid, HI	1·98	1·90
Nitric acid, HNO_3	2·03	1·96
Sulphuric acid, H_2SO_4	1·84	1·93
Acetic acid, CH_3COOH	1·03	1·01
Lactic acid, $C_3H_6O_3$	1·01	1·03
Barium chloride, $BaCl_2$	2·63	2·54
Potassium chloride, KCl	1·86	1·86
Sodium chloride, $NaCl$	1·90	1·82
Ammonium chloride, NH_4Cl	1·88	1·84
Mercuric chloride, $HgCl_2$	1·11	1·05
Potassium iodide, KI	1·90	1·92
Potassium nitrate, KNO_3	1·67	1·81
Sodium nitrate, $NaNO_3$	1·82	1·82
Sodium carbonate, Na_2CO_3	2·18	2·22

The conclusion drawn by Arrhenius was that the reason for the apparent anomaly in the osmotic pressures and depressions of freezing-point in solutions of electrolytes in water is that the molecules of the electrolytes are dissociated to a greater or less extent into ions, so that the true molecular concentration is in reality greater than corresponds to the theory that the molecules are present in the undissociated state.

The theory of Arrhenius appeared at the time so revolutionary that many contemporary chemists were unable to accept it. But confirmation came from various other directions; and all doubts as to the truth of the theory have long disappeared.

It was supposed at first that when the molecules of an electrolyte are ionized, the atoms or atomic groupings into which the original molecules split up are present in the solution in an uncombined form, so that, for instance, when a potassium chloride molecule is ionized, isolated, but electrically charged, atoms of potassium and chlorine pass into the solution; or that when a molecule of potassium sulphate is ionized two isolated ions of potassium and one of SO_4 are formed. Evidence has, however, accumulated that the ions are combinations of the fragments of the original molecule with molecules of water. In other words, the ions are heavily hydrated, and may be regarded as a product of the splitting up of the original molecules by the action of water molecules which have entered into combination with ions, and so torn them apart from their combination in un-ionized molecules.

This has an important bearing on the osmotic pressure (and depression of freezing-point, etc.) of aqueous solutions of electrolytes. We have seen how important the existence of hydration is in influencing the osmotic pressure given by sugar solutions; and there can be no doubt as to a similar influence of hydration in the case of the ions, as well as the un-ionized molecules, of solutions of electrolytes. Existing information on this influence is still very incomplete; and unfortunately the mistaken theory of van't Hoff that osmotic pressure, etc., depend on the concentration of solute molecules per unit volume of solution has directed investigation away from this question, since on van't Hoff's theory it would not matter, so far as osmotic pressure, etc., are concerned, whether or not the ions are hydrated.

When an ionized solute is present, we have to consider the question whether or not each of the ions into which the dissolved substance breaks up is capable of penetrating a given semi-permeable membrane. If one sort of ion can penetrate, but the other cannot, a somewhat complex situation arises. The two sorts of ions into which the molecules break up are of opposite electrical charges. If, therefore, one sort passes out through a membrane, and the other does not, a difference of charge on the two sides of the membrane will quickly be produced, and this will tend to stop the diffusion through the membrane of the ion which can penetrate. We have thus differences of electrical charge complicating the phenomena of osmotic pressure. The quantitative relations of this phenomenon were worked out by Donnan[1]; and Loeb has recently shown in detail how important these relations are in connection with many of the phenomena observed in solutions of colloids.[2]

The error in all the earlier experiments of Morse on the osmotic pressures of sugar solutions (including dextrose solutions) was probably due to complications produced by permeability of the copper ferrocyanide membrane by the SO_4 ions on the outside, but not by the potassium ions inside the cell. Differences of potential between the inside and outside of the cell would thus arise, and the osmotic equilibrium which would exist if the membrane were impermeable to all the ions would not be present. There would thus be an error of a certain fixed amount in the osmotic pressure attributed to the sugar solution, and at the lower concentrations of sugar this error would be serious, as was actually the case. At a sugar concentration (weight-normal) ten times as great the percentage error (corrected for the error of the nitrogen gauge) would only be a tenth as great, as was also actually the case.

It is because of the permeability of copper ferrocyanide membranes and the external layers of vegetable or animal protoplasm to the ions of one electrical charge that these membranes cannot be used for determining the osmotic pressures which would otherwise be given by solutions of electrolytes. Hamburger, for instance, has shown that though red blood corpuscles can be used, just as de Vries used vegetable cells, for comparing different solutions as regards their isotonicity, and give the same results, yet the walls of the corpuscles are fairly easily permeable to chlorine ions or SO_4 ions, etc. On the other hand, they are impermeable to sodium or potassium ions. In some animals the red corpuscles contain much potassium but practically no sodium, though the plasma in which they pass their existence contains much sodium and very little potassium. The impermeability or relatively slow permeability of the surface layer of living cells for various ions or molecules furnishes an explanation of a number of otherwise anomalous physiological facts. It has also been shown in many cases that the conditions which determine impermeability or permeability are under physiological control.

---

[1] Donnan, *Z. f. Elektrochemie*, xvii., p. 572, 1911.

[2] Loeb, *Proteins and the Theory of Colloidal Behaviour*, 1922.

Evidence will be referred to below (Chapter VII.) that in the case of water and other solvents the molecules are at any moment to a large extent aggregated together as minute but quite unstable solid particles. The question therefore arises whether this aggregation will have any effect on osmotic pressure or depression. The aggregates may, as in the case of water, occupy more virtual space than would the corresponding unaggregated molecules: or they may occupy less space. To realise the effect of aggregation of solvent molecules we have to consider that the expression for osmotic pressure is $\frac{nN}{N_1 - n} .082\,T$. Of the terms in this expression, N, which expresses the molar concentration of pure solvent, and $N_1 - n$, which expresses molar concentration of the solvent in the solution, are affected by the aggregation. But if these two terms are affected in equal proportion the effect of the aggregation on osmotic pressure cancels out, since N is a numerator and $N_1 - n$ the denominator in the whole expression. The actual figures for osmotic pressures, etc., with aqueous solutions show clearly that these two terms are, in fact, affected in equal proportions. Hence, in calculating osmotic pressure we need take no account of aggregation of solvent molecules. We can thus see why, as a matter of fact, the results calculated by the equation work out correctly in spite of the large extent of aggregation among the molecules of water.

Aggregation of solute molecules, or of solute with solvent molecules, is a very different matter, as we can see from the equation, and as we have already seen from the actual influence of hydration of solute molecules in affecting osmotic pressure.

It seems probable that osmosis, or diffusion of water, is the means by which ordinary spontaneous movements are brought about in plants and animals. As regards plants, there is much direct evidence on this point. The soft parts of plants consist of cells with an outer cellulose wall which is permeable to water and dissolved substances. Inside this outer wall is living protoplasm, and it is evident that this protoplasm has usually absorbed so much water as to keep the outer wall in a state of tension and prevent its collapse. Collapse ensues, however, when the plant is deprived of its supply of water, or when it is supplied, in place of ordinary water, with a solution of considerable osmotic deficiency—a fairly strong sugar solution, for instance. The soft parts collapse, droop, and "wilt." The method of de Vries for measuring isotonicity of solutions (page 134) depends on the observation of this wilting process through the microscope. The cells in the soft parts of normal plants are normally blown out with water; and it seems evident that the reason for this is that osmotic depression is greater inside the protoplasm than in the sap. In animals there is exactly the same normal blowing out of the tissue elements with water as in plants.

In so-called sensitive plants the protoplasm of certain groups of cells has the power of rapidly discharging water and so letting the cell-walls collapse. This discharge was observed directly by Pfeffer, and it brings about various bending or contracting movements. A rapid diminution of osmotic depression in the protoplasm appears to be the cause of the discharge of water. In the case of other protoplasmic movements, as seen through the microscope, local changes in osmotic depression seem to be much the most probable immediate cause of movement.

The immediate cause of muscular movement has for long been a subject for speculation. Descartes suggested that muscles contract because cavities in them are blown out by a fluid conveyed by the nerves; but this theory was evidently incorrect, and subsequent investigators have been somewhat at a loss for a theory as to how muscles contract. There seems, however, to be no reason for believing that there is any fundamental difference in the causes of animal and plant movements, and, as Burdon Sanderson showed, similar electrical changes accompany them. Osmotic changes provide an abundance of driving-power for contractile movements; and the mechanical energy may be obtained at the *immediate* expense of the molecular energy which we call heat, the energy of this movement being so harnessed in as to be effective. (See Chapter VIII., page 328.)

At first sight it may seem difficult to understand how local differences in osmotic depression can be produced in living protoplasm. There is now, however, abundant evidence that the molecules of colloid albuminous solutions as they exist in living protoplasm possess the property of entering into easily dissociable combinations with other smaller molecules in response to very minute changes in their physical environment. This process must have the effect of altering the osmotic depression, and so leading to forcible migration of water.

Contractile tissues have a structural arrangement which suggests that osmotic or other locally produced pressure effects act in one direction only. Thus a voluntary muscle is a large bundle of similarly disposed muscle-fibres; and each muscle-fibre is a bundle containing a very large number of muscle-fibrils, all arranged similarly. The theory was put forward a few years ago by Professor W. McDougall[1] that muscular contraction is due to the imbibition

---

[1] McDougall, *Journ. of Anat. and Physiol.*, xxxi., p. 410, 1896.

of water by these fibrils, owing to changes in their reaction when they are excited by a suitable stimulus. On this theory they are blown out by the imbibed water, and therefore take a more spherical form, with consequent shortening of the fibril. McDougall's theory was thus somewhat similar to that of Descartes, though the distending substance was supposed to be simply water, instead of the hypothetical "animal spirits" postulated by Descartes. It seems much more probable, however, that the movement of water is due to osmotic change owing to dissociation of smaller molecules or ions from protein molecules, and consists in such a migration of water within each segment of the fibril that owing to the mechanical structure of the fibril it becomes shorter and at the same time thicker.

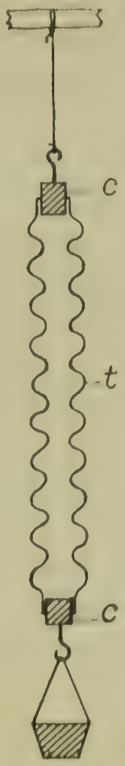

FIG. 6.—Model to illustrate Osmotic Contraction.

The corrugated india-rubber tubing *t* is filled with carbon dioxide and is closed with corks *c*, *c*, above and below and hung up with a weight attached below. The tube slowly contracts and lifts the weight, since the india-rubber is freely permeable to carbon dioxide, but not to air.

To illustrate a manner in which osmotic changes may cause contraction I arranged the above working model.[1] A piece of corrugated rubber tubing of the kind devised by Fleuss for breathing through in connection with his mine-rescue apparatus was filled with carbon dioxide, closed at both ends, and hung up with a weight attached below (Fig. 6). As carbon dioxide is far more soluble in rubber than air is, the carbon dioxide passes out through the rubber far more quickly than air passes in. The rubber thus plays the part, as a semi-permeable membrane, of the surface of a protoplasmic layer. The tube gradually contracts and lifts the weight, till at last the corrugations of the tube are firmly pressed together and extended sideways, if the weight is not too heavy. With the model several hours are required for the full contraction; but an easy calculation shows that if the model were on the scale of a muscle-fibril the contraction would occur in less than a thousandth of a second.

A further discussion of this subject will be found at the end of Chapter VIII.

---

[1] This model was shown contracting at a meeting of the Physiological Society in 1922; but a description of it has not been published till now.

CHAPTER V

DEPRESSION OF FREEZING-POINTS AND VAPOUR PRESSURES, AND ELEVATION OF BOILING-POINTS, OF SOLUTIONS

It has been known since Blagden's observations in the eighteenth century that when a substance is dissolved in water the freezing- or melting-point of the water is lowered, and the more so the stronger the solution of any particular substance. Similarly, the boiling-point is raised. For example, the boiling-point of a solution containing 25 grammes of potassium chloride per 100 grammes of water is raised to 103.44° at atmospheric pressure, while its freezing-point is lowered to – 10.9°. The raising of the boiling-point shows, of course, that the vapour pressure of water is lowered by the presence of the solute.

These phenomena in connection with solutions, not only in water, but also in other solvents, were carefully investigated by Raoult,[1] who made the striking discovery that for any solvent the lowering of melting- or freezing-point is, in the case of a large variety of organic solutes, proportional to the number of moles dissolved in a given weight, such as 100 grammes, of *solvent* (not in a given weight or volume of *solution*). On this account observations of depression of freezing-point became of the utmost value in determining the molecular weights of organic substances, and are now universally used for this purpose. Hence if, adhering to the nomenclature adopted above, we denote by n the number of moles of solute in a litre of solution, by N and N_1 the number of moles respectively in a litre of pure solvent and of solution, and by S the specific gravity of the pure liquid solvent, at the freezing- or melting-point, the depression of freezing-point is by Raoult's law proportional to $\frac{nN}{(N_1 - n)S}$, or to moles of solute per 1000 grammes of solvent (see p. 112), or, since $\frac{N}{S}$ is constant, to $\frac{n}{N_1 - n}$, just as in the case of osmotic pressure at the freezing-point.

---

[1] Raoult, *Annales de Chim. et de Phys.*, (5) xxviii. 1883; and (6) viii. 1886.

The depression is not simply proportional to n, the molar concentration of the solute, although at very great molar dilutions $\frac{n}{N_1 - n}$ varies almost exactly as n since $N_1 - n$ becomes almost equal to the constant N. In the case of such solutes as colloids, which possess very large molecules, the molar dilution has to be enormous before practical equality is reached. On the other hand, where solute and solvent molecules are nearly equal in size, the values of $N_1 - n$ and N diverge only slowly as the molar concentration increases. In the case of aqueous solutions at temperatures not far from 4°, S = 1.000, and may therefore be neglected.

Raoult found that for each solvent there is a constant which, when multiplied by the number of moles of solute per 100 grammes of solvent, gives the number of degrees by which the freezing-point is lowered. When, for instance, water is the solvent, and the concentration is reckoned, as Raoult reckoned it, in moles of solute per 100 grammes of solvent, this constant is 18.7, or 1.87 if the concentration is reckoned in moles per 1000 grammes of solvent. He named the constant for each solvent its cryoscopic constant (from κρύος, icy cold, and σκοπέω, I look). This constant is also known as the "molecular lowering of freezing-point"; but the latter expression is rather apt to be misleading, since it suggests that the lowering is simply proportional to the molecular or molar concentration of the solute (*i.e.* to *n*).

Raoult did not realise that in determining the cryoscopic constant it is necessary to take account of the formation of hydrates or other combinations ("solvates," as H. C. Jones called them) between solute and solvent; but when sufficiently dilute solutions are used the influence of solvates on determinations of the cryoscopic constant is only small. Until Arrhenius published the famous paper referred to in Chapter II., Raoult was also unaware of the influence of ionization in solutions of electrolytes. He was thus unable to account for the fact that in the case of electrolytes the cryoscopic constant, as deduced in the ordinary way, was quite abnormally high, and varied with the strength of the solution and nature of the solute. The discovery of Arrhenius made it possible, however, to apply Raoult's generalisations, not only to organic solutes, but also to all other solutes, though the complications due to hydrates or other solvates, and to association with one another of solute molecules, still remained.

The following table shows how accurately Raoult's law applies in the case of cane-sugar solutions where the allowance required by the hydration of the cane sugar is made. The experimental data used in constructing this table are those quoted from various observers in the third edition of Landolt and Börnstein's *Physikalisch-chemische Tabellen*, except in the case of the last four determinations, which are from Jones and Getman.[1] Of the three columns giving the calculated cryoscopic constant, the first is calculated on the assumption that the sugar is present as pentahydrate, in accordance with the corresponding

Depressions of Freezing-Point of Cane-Sugar Solutions.

Grammes Anhydrous Sucrose per 100 gm. of Total Water.	Moles Sucrose Pentahydrate per 1000 gm. Free Water.	Observed Depression of Freezing-Point.	Cryoscopic Constant.		
			Calculated for Sucrose Pentahydrate from Columns 2 and 3.	Calculated for Anhydrous Sucrose from Columns 1 and 3 and Molecular Weight. (Morse's Principle.)	Calculated from Molar Concentration deduced from Column 1, Specific Gravity, and Molecular Weight. (Van't Hoff's Principle.)
·04825	·001409	·00264°	1·874°	1·87°	1·87°
·3414	·00999	·01856°	1·858°	1·86°	1·86°
·6875	·02014	·0378°	1·877°	1·88°	1·89°
2·231	·0656	·1230°	1·875°	1·89°	1·91°
4·465	·1320	·2450°	1·856°	1·88°	1·93°
29·82	·944	1·768°	1·873°	2·03°	2·34°
56·18	1·923	3·630°	1·888°	2·21°	2·98°
68·90	2·461	4·612°	1·874°	2·29°	3·29°
84·46	3·178	5·800°	1·825°	2·35°	3·62°
101·99	4·077	7·230°	1·77°	2·43°	4·02°
122·14	5·27	9·13°	1·72°	2·54°	4·57°

---

[1] Jones and Getman, *Amer. Chem. Jour.*, 32, p. 308, 1904.

calculated osmotic pressure in the table on page 149. The second is calculated on the assumption that the sugar is not hydrated; while the third is calculated from the theory of van't Hoff that osmotic pressure, and consequently depression of freezing-point, is in dilute solutions proportional to the number of moles of solute present per unit volume of solution.

It will be seen that at all strengths of solution until the proportion by weight of free water is less than half of that of sucrose pentahydrate the first column gives a constant of 1.87°. At higher concentrations the calculation begins to give too low a result, just as in the case of osmotic pressure, and presumably for the same reason. The second and third of the columns give about the same figure as the first when the concentration of the sugar is less than 0.5 per cent., and the actual depression of freezing-point less than 0.02°. At higher concentrations the calculated figures increase, slowly in the case of the second last column, but much less slowly for the last. At the highest concentration the results are about 36 per cent. too high for the second last, and 144 per cent. too high for the last column. The table seems to leave no doubt, not only as to the correctness of Raoult's method of calculating the cryoscopic constant, or "molar" or "molecular" depression of freezing-point, but also as to the necessity for taking hydration into account in the calculation.

Very numerous observations by other observers have confirmed for dilute solutions Raoult's conclusion that depression of freezing-point of the solvent varies with the number of moles of solute to a given weight, or a given number of moles, of solvent in the solution; and it is not necessary to cite here further examples of the law. Data from Arrhenius showing that in the case of solutions of electrolytes Raoult's law holds good if ionization of the solute molecules is allowed for, have already been given at the end of Chapter IV.

In the paper already referred to van't Hoff pointed out that it follows, on thermodynamical principles, that if osmotic pressure varies with molar concentration per unit volume of solution, depression of freezing-point must do so likewise. When a solvent crystallises, this is because its temperature is so low that more molecules pass into the solid form than pass out of it into the liquid form. In other words, the heat-energy which promotes dissociation of the molecules is insufficient to maintain the solvent in the liquid state. According to van't Hoff's conception, osmotic pressure represented a positive influence which drives solvent into a solution, and which varies with the molecular concentration, n, of the solution. It must therefore lower the freezing-point, since it acts against the influences which promote separation out of the solvent molecules; and since temperature is simply a measure of molecular kinetic energy, the depression in the freezing-point of a solution must be proportional to its osmotic pressure, and must therefore vary, like osmotic pressure, with the concentration of the solute. There would be no flaw in van't Hoff's inference that depression of freezing-point must vary as the molar concentration of the solute if it were true that the osmotic pressure varies as the molar concentration of the solute. But this, as we have seen, is not true.

Van't Hoff's reasoning is independent of any theory as to the cause of osmotic pressure; and this has led to its general acceptance, even by those who were not convinced by his explanation of the cause of osmotic pressure. At the time when he wrote his paper the only quantitative data as to osmotic pressure were those of Pfeffer; and they supported his assumption that osmotic pressure varies with the molar concentration of the solute. On the other hand, the experimentally ascertained cryoscopic constants were only consistent with this assumption in the case of solutions so dilute that the difference between molar concentration per unit volume and the number of moles of solute to moles of solvent per unit volume of solvent became inappreciable. De Vries had, however, already found by the plasmolysis method that osmotic equivalence between solutions seemed to vary with Raoult's ratio of moles of solute to weight of solvent.

Van't Hoff smoothed over the inconsistency between the facts and his assumption by arguing that in the case of the less dilute solutions it was to be expected that his assumption would not hold, since even with gases it is only in the case of great molar dilution that the equation $PV = RT$ holds exactly, and this must also be anticipated for solutions. In actual fact, as we have seen, the equation $PV = RT$ is not strictly true at any finite dilution of a gas, and cannot be even remotely near to being true for a gas in a mixture of the very high molar concentration of a solution. But even if it were true as applied to the bombardment pressures of solute and solvent, it would give an osmotic pressure and depression of freezing-point varying, not with the molar concentration of the solute per unit volume, but with Raoult's ratio of moles of solute to weight (or moles) of pure solvent (see page 111).

The general acceptance of van't Hoff's explanation of the inconsistency has been most unfortunate in its effects on the further developments of physical chemistry. It is owing to this acceptance that our data as to strong, or even moderately strong, solutions are still so scanty, and that the influence of hydration or solvation of ions and undecomposed solute molecules, and of association of solute molecules has not been properly worked out as yet. A great field for experimental research is still almost unreaped.

It was shown in Chapter IV. that van't Hoff's assumption that osmotic pressure varies as the concentration of solute molecules, and his theory of the cause of osmotic pressure, were incorrect. If we substitute for his assumption the correct assumption that osmotic pressure varies as the ratio of moles of solute to weight, or moles, of solvent, the essential element in his brilliant thermodynamical reasoning remains. If, moreover, we substitute for his theory of the cause of osmotic pressure the theory which has been developed in the previous chapters, all the facts are brought into definite relation with the fundamental gas laws, provided they are stated in the form given to them on page 20.

Quite apart from the details of his theory of osmotic pressure, van't Hoff made the extremely important, and quite evident, thermodynamical deduction that if the osmotic pressure in atmospheres equals $n\,.082\,T$, which may be expressed in calories $n\,1.983\,T$ (see p. 24), and the latent heat of melting one gramme of solvent is L calories, the depression of freezing-point is $\frac{n\,1.983\,T^2}{1000\,L}$, which may be simplified into the form $\frac{n\,.002\,T^2}{L}$, (or $\frac{n\,.02\,T^2}{L}$ if n is taken as moles per 100 c.c.), as being sufficiently close. We shall return later to the thermodynamical reasoning by which van't Hoff reached this conclusion. Meanwhile, however, it is sufficient to point out that if for n in this expression we substitute $\frac{nN}{(N_1 - n)S}$ which expresses a concentration with which both osmotic pressure and depression of freezing-point actually vary, van't Hoff's expression gives with great accuracy the cryoscopic constants discovered by Raoult for different solvents, or, alternately, gives from actual determinations of depressions of freezing-point the latent heat of melting of the solvent.

As an example of the accuracy with which the calculated cryoscopic constant agrees with experiment, when due allowance is made for hydration, the data in the table on p. 178 may be referred to. The calculated value of $\frac{1.983\,T^2}{1000\,L}$, since $T = 273.1°$ and $L = 79.3$, is 1.87°. It will be seen from the table that until the extremely high concentrations at which alteration in internal pressure presumably begins are reached, no definite deviation from the value 1.87° appears in the values calculated from actual experiments, though the data are from different observers, not one of whom was aware of the correct method of calculating the cryoscopic constant, which is for sucrose pentahydrate. Examples for other solvents, though without the corrections required for great exactitude, were given by van't Hoff and subsequent observers, and may be found in text-books, but much more completely in Landolt and Börnstein's *Tabellen*.

Not only are certain solutes (electrolytes) dissociated more or less completely into ions on solution, thus increasing the apparent molar depression, but in the case of certain other solutes some or all of the molecules are evidently present in combination with one another, particularly when the solvent is an organic one. This, of course, diminishes the apparent molar depression correspondingly.

We may now turn to the vapour pressures of solutions. In a series of observations on the depressions produced in the vapour pressure of a solvent by the presence of various organic solutes, Raoult[1] made the discovery that up to considerable molar concentrations of the solute, and to very easily measurable values of vapour pressure, the percentage depression varies, in spite of variation in temperature, as the ratio of moles of solute to the total moles in the solution (not as the moles of solute per unit volume or weight of either solution or solvent). This ratio corresponds to $\frac{n}{N_1}$ in the notation adopted in this book. Hence if we call p the vapour pressure of the pure solvent, and p_d the depression of vapour pressure, $p_d = \frac{n}{N_1}p$, or $\frac{p_d}{p} = \frac{n}{N_1}$, or, calling p_1 the vapour pressure of the solvent in the solution, $\frac{p-p_1}{p} = \frac{n}{N_1}$.

---

[1] Raoult, *Compt. rend.*, 22, July 1878; *Z. f. physik. Chem.*, ii., p. 353, 1888.

The table on page 186, calculated from Raoult's data, illustrates this law in the case of several solutes, ether being chosen as the solvent since it has an easily-measured vapour pressure, while the solutes have such low vapour pressures as not to disturb the measurements appreciably. As was pointed out in Chapter II., the osmotic pressure of a solution varies as $\frac{n}{N_1-n}$, or moles of solute to moles of solvent. A separate column in the table is therefore added in order to show how the depression of vapour pressure varies in relation to $\frac{n}{N_1-n}$.

On examining the table and allowing for probable errors of experiment, it will be seen that $\frac{p_d}{p}$ varies as $\frac{n}{N_1}$, and not as $\frac{n}{N_1-n}$. With stronger solutions, the results for which will be discussed in Chapter VII., the value of $\frac{p_d}{p}$ becomes less than that of $\frac{n}{N_1}$, but a mere fraction of that of $\frac{n}{N_1 - n}$. It is evident, indeed, that the limiting value of both $\frac{p_d}{p}$ and $\frac{n}{N_1}$ is 1, while the limiting value of $\frac{n}{N_1 - n}$ is infinity, and that of $\frac{n}{N}$ may be anything, according as the relative sizes of solute and solvent molecules vary.

Depressions of Vapour Pressure in Solutions in Ether.

(Molecular weight of Ether, $C_4H_{10}O = 74$. Boiling-point = 34·6°.)

Grammes of Solute per 100 g. of Solution.	Moles of Solute to Moles of Solvent. $\frac{n}{N_1 - n}$	Moles of Solute to Moles of Solution. $\frac{n}{N_1}$	Relative Depression of Vapour Pressure of Ether. $\frac{p_d}{p}$
(1) Oil of Turpentine, $C_{10}H_{16}$. M.w. = 136. B.p. = 160°			
10·0	·060	·057	·060
20·2	·138	·121	·119
(2) Nitrobenzol, $C_6H_5NO_2$. M.w. = 123. B.p. = 205°			
9·6	·064	·060	·055
(3) Anilin, C_6H_7N. M.w. = 93. B.p. = 182°			
4·8	·040	·039	·040
9·5	·084	·077	·081
18·1	·176	·150	·154
(4) Methyl Salicylate, $C_8H_8O_3$. M.w. = 152. B.p. = 222°			
4·2	·022	·021	·017
9·4	·051	·048	·040
17·3	·101	·092	·086
(5) Ethyl Benzoate, $C_9H_{10}O_2$. M.w. = 150. B.p. = 213°			
9·4	·051	·049	·051
17·7	·106	·096	·091

For the depression of vapour pressure of water in cane-sugar solutions at 0° there exists a valuable series of observations by Smits.[1] From these observations I have calculated the following table, assuming that the sugar is present as pentahydrate.

Depressions of Vapour Pressure of Cane-Sugar Solutions at 0° C.

(Molecular Weight of Sucrose Pentahydrate, $C_{12}H_{22}O_{11} + 5H_2O = 432{\cdot}1$)

Grammes Anhydrous Sucrose per 100 g. Total Water.	Moles Sucrose Pentahydrate to Moles of Free Water. $\frac{n}{N_1 - n}$	Moles Sucrose Pentahydrate to Moles of Solution. $\frac{n}{N_1}$	Relative Depression of Vapour Pressure of Water as observed. $\frac{p_d}{p}$	Moles Sucrose per litre of Solution to Moles Water per litre of Pure Water. $\frac{n}{N}$
1·711	·000905	·000904	·00091*	·00087
5·896	·00315	·00314	·00320	·00291
15·54	·00853	·00846	·00846	·00730
34·52	·0200	·0196	·0196	·0144

* In Smits' paper there is an evident misprint, copied in Landolt and Börnstein's *Tabellen*, of 0·00242 for 0·0042. In my calculation this is allowed for.

The molecular weight of sucrose pentahydrate is so high, and of water so low, in comparison with the solutes and solvent used by Raoult, that the difference between columns 2 and 3 in this table is very small. Nevertheless, it will be seen that the ratio $\frac{p_d}{p}$ agrees more closely with the ratio $\frac{n}{N_1}$ than with the ratio $\frac{n}{N_1 - n}$. Raoult's law therefore holds for aqueous solutions. On comparing columns 4 and 5 it will be seen at once that depression of vapour pressure does not run parallel with molar concentration of solute.

---

[1] Smits, *Zeitschr. f. pysikal. Chemie*, 51, p. 33, 1905.

Berkeley, Hartley, and Burton have determined the depressions of vapour pressure in very strong solutions of cane-sugar. Their results will be discussed in Chapter VII.

The actual measurements of Raoult and others have often been ignored, and are even grossly misrepresented in some text-books. According to van't Hoff's conception the ratio $\frac{p_d}{p}$ ought to equal $\frac{n}{N}$. The last column in the table shows in the clearest manner that this is not the case. This column is calculated from column 1 and the known specific gravities of cane-sugar solutions at 0°. A similar column could not be calculated for Raoult's results, as the specific gravity data are not available; but evidently the value of $\frac{n}{N_1}$ would, in the case of Raoult's experiments, diverge increasingly from that of $\frac{n}{N}$ as the concentration increased. If, however, the molecular weights of solute and solvent were equal, the values of $\frac{n}{N_1}$ and $\frac{n}{N}$ might never diverge.

There exist also some important observations in which Berkeley, Hartley and Burton compared the osmotic pressures and depressions of vapour pressure in strong solutions of calcium ferrocyanide.[1] Their results are given in the table on opposite page.

---

[1] Berkeley, Hartley and Burton, *Phil. Trans.*, A, 209, p. 177.

Depressions of Vapour Pressure and Osmotic Pressures of Calcium Ferrocyanide Solutions at 0° C.

Grammes Anhydrous Calcium Ferrocyanide per 100 g. Total Water.	Osmotic Pressure in Atmospheres.	Percentage Depression of Vapour Pressure.	Relative Values.	
			Osmotic Pressure.	Depression of Vapour Pressure.
31·388	41·22	3·210	1·000	1·000
39·503	70·84	5·400	1·719	1·682
42·889	87·09	6·557	2·113	2·042
47·219	112·84	8·397	2·737	2·622
49·996	130·66	9·674	3·170	3·014

It will be seen from this table that the depression of vapour pressure does not increase as rapidly as the osmotic pressure. The values of $\frac{n}{N_1}$ and $\frac{n}{N_1 - n}$ cannot, however, be calculated directly from the data.

The value of the actual depression of vapour pressure of solvent in any solution depends, of course, on the temperature, since the vapour pressure increases with the temperature. Raoult showed, however, that the ratio $\frac{p_d}{p}$ for a solution is independent of variations in the temperature of the solution. The next table (p. 190) illustrates this independence in the case of solutions in ether.

The vapour pressure of a substance is evidently the pressure of its vapour at which inward diffusion balances the outward diffusion consisting in the flying off of molecules from its surface. In accordance with the conception of a liquid as a system of freely moving molecules held in proximity to one another by molecular attraction, but subject otherwise to the fundamental gas laws, the escape of molecules from the surface is due to the translational kinetic energy of some of them being for the moment so great in the direction at right angles to the surface of the liquid that they fly off from the surface, thus producing external pressure in a closed vessel. By the law of probability as applied to enormous numbers of molecules this proportion must be constant at any given temperature, and must vary inversely as the excess, over the average translational energy of the molecules in the liquid, of the translational energy needed to carry the molecules free. This excess, as was pointed out by van't Hoff, and was long previously indicated by Waterston, is just, per gramme of liquid, the latent heat of evaporation

RELATIVE DEPRESSIONS OF VAPOUR PRESSURE AT DIFFERENT TEMPERATURES.

1. 16.48 *grm. Oil of Turpentine to* 100 *grm. Ether.*

Temp. C.	p for Pure Ether	$p+p_d$	$\frac{p-p_d}{p}\times 100$
1·1°	199·0	182·1	91·5
3·6°	224·0	204·7	91·4
18·2°	408·5	368·7	91·0
21·8°	472·3	430·7	91·2
2. 10.42 *grm. Anilin to* 100 *grm. Ether.*			
1·0°	199·5	183·1	91·9
3·6°	223·2	204·5	91·6
9·9°	289·1	264·0	91·3
21·8°	472·9	432·7	91·5

at the existing temperature. The proportion, per unit of free surface, must also depend on the molar concentration, and be connected in some way with temperature, since the latter, in accordance with the dynamical theory, expresses the mean translational energy of the molecules in the liquid. Further discussion of these points in connection with Waterston's experiments and reasoning will be found at page 279.

Since at anywhere near or below ordinary atmospheric pressure the value of $1 - Nb$ is practically the same in a gas or vapour as 1, at such a pressure the rate of escape of molecules from the surface will vary inversely with the excess of kinetic energy required in the molecules about to escape, as there is no appreciable hindrance from the atmosphere. In accordance, however, with the principles laid down in connection with osmotic phenomena, the rate of passage of vapour molecules from the atmosphere into the liquid must vary as the pressure of the vapour molecules and as the value of $1 - N_1b$ for the particular liquid, just as in the case of liquid diffusion. Hence, if the vapour pressure is p,[1] the inward diffusion pressure which balances the outward is $(1 - N_1b)p$, and not p. The diffusion pressure of a vapour or gas into or out of a liquid is thus, at about atmospheric pressure, only a fraction of the total vapour or gas pressure, just as the diffusion pressure of one liquid into another is only a fraction of its internal bombardment pressure.

---

[1] At temperatures where the molar concentration of the vapour becomes considerable the external pressure of the vapour no longer represents its total pressure, and as will be shown below this must be taken into account in certain calculations.

If the surface of the liquid is partly covered by a layer impermeable to the molecules of the liquid, or if there is surface condensation of solute molecules, this will not affect the vapour pressure, since outward and inward diffusion are equally affected. Nor will the altered concentration at the plane surface of any liquid affect the vapour pressure.

If, now, a solvent is diluted by a solute with no appreciable vapour pressure and the internal bombardment pressure of the liquid is not thereby altered, and the temperature is constant, the pressure of vapour molecules outwards will necessarily be diminished (as shown in Chapters I. and III., pp. 25 and 111) in the proportion of 1 to $\frac{N_1 - n}{N}$, where n, N, and N_1 denote respectively moles of solute, pure solvent and solution per litre. As already seen in the previous chapters, the values $1 - Nb$ for the pure solvent, and $1 - N_1 b_1$ for the solution vary as N to N_1, so long as the internal bombardment pressure P is the same in pure solvent and solution, as was inferred in connection with osmotic pressure to be the case up to at least very considerable concentrations of solute. Hence, if p be the vapour pressure of the pure solvent the diffusion pressure of the solvent outwards from the solution will be diminished to $\frac{N_1 - n}{N}(1 - Nb)p$, which equals $\frac{N_1}{N}(1 - Nb)p - \frac{n}{N}(1 - Nb)p$; and the diffusion inwards, if P remained the same, would become $\frac{N_1}{N}(1 - Nb)p$. Hence the diffusion inwards will be greater by $\frac{n}{N}(1 - Nb)p$. Vapour will therefore condense, so that the vapour pressure will fall and a lower vapour pressure will establish itself. If the diminution of vapour pressure is designated as p_d, $\frac{N_1 - n}{N}(1 - Nb)p$ must be equal to $\frac{N_1}{N}(1 - Nb)(p - p_d)$. From this equation it follows that $\frac{n}{N}p = \frac{N_1}{N}p_d$, so that $\frac{p_d}{p} = \frac{n}{N_1}$. In other words, the relative depression of vapour pressure of the solvent is equal to the ratio of molar concentration of the solute to molar concentration of the solution.

Now this, as we have seen, is just what was discovered experimentally by Raoult. Hence the same assumption (that the internal bombardment pressure remains equal in the solution and the pure solvent) which enables us to predict variations of osmotic pressure enables us also to predict variations in depression of vapour pressure in solutions, up to very considerable concentrations.

By similar reasoning it can be shown that where the solute has a measurable vapour pressure this pressure will vary directly as $\frac{n}{N_1}$, or the ratio of moles of solute to total moles per litre of solution. We can also, instead of speaking of depression of vapour pressure of solvent, say alternatively that when a solvent is diluted up to a limited extent by a solute the vapour pressure of the solvent varies with $\frac{N_1 - n}{N_1}$, the ratio of moles of solvent to total moles in the solution. As $\frac{N_1 - n}{N_1}$ equals $1 - \frac{n}{N_1}$ this alternative statement is identical with Raoult's statement.

It was shown in Chapter IV. (p. 160) that the deficiency or depression in the osmotic pressure capable of being exercised by any solution in a given solvent, as compared with that capable of being exercised by the pure solvent, varies within wide limits of concentration as the value of $\frac{n}{N_1}$ for the solution, and is given by the expression $\frac{n}{N_1}(1 - Nb)P$, or $\frac{nN}{N_1}$ .082 T. This is the osmotic depression P_d, which exists for every solution, and which is abolished when the external pressure P_o (*i.e.* the osmotic pressure) is applied to the solution. Hence P_d, the osmotic depression, varies as p_d, the depression of vapour pressure. Thus $\frac{P_d}{P} = \frac{p_d}{p}$, and just as P_d is to P as n to N_1, so p_d is to p as n to N_1. The great importance of this conclusion in physiology was pointed out in my recently-published book *Respiration.*[1]

The surface of a liquid is seldom sensibly impermeable to the passage outwards of molecules of the solvent, and the vapour pressure of the latter is the pressure at which diffusion equilibrium is established between escaping and returning molecules. Hence, if the solute has no appreciable vapour pressure the surface behaves, to the extent of the outward pressure of the escaping molecules, like a membrane permeable only to the solvent. The physical conditions determining evaporation from a liquid surface will be further discussed in Chapter VII.

---

[1] Haldane, *Respiration*, Yale University Press and Oxford University Press, 1922, p. 383.

If, instead of allowing the pressure of the vapour in contact with a solution to fall below that of the pure solvent, we raise the pressure on the solution until its vapour pressure is the same as that of the pure solvent, the increase of pressure required is, if calculated on the same principles as those on which the depression of vapour pressure was calculated, $\frac{nN}{N_1 - n} .082\,T$, which is the osmotic pressure. By confining the liquid in the rigid cell of an osmometer, we can put pressure on the liquid without thereby putting pressure on the atmosphere. If the outside of the cell was dried as far as possible, and the surrounding air kept saturated with vapour from pure solvent lying on the bottom of the external vessel, the experiment could be made directly. But if the cell were dipped in pure solvent this could not affect the result, since if, for instance, we imagined that solvent in the form of vapour could penetrate the cell, while the pure solvent could not do so, the result of the penetration would be that the osmotic equilibrium would be disturbed, and the pure solvent would pass out again. If, on the other hand, we imagined that the vapour of the solvent passed out into the atmosphere, though pure solvent itself could not do so, there would also be a disturbance of the equilibrium, and the pores in the external layers of the membrane would be dried up, which would reverse the direction in which the vapour was supposed to be passing.

Since with increase in the value of n the value of $\frac{n}{N_1 - n}$ increases more rapidly than the value of $\frac{n}{N_1}$, it is clear that osmotic pressure and relative depression of vapour pressure do not run parallel to one another in solutions of varying molar concentration of solute, and this is clearly shown from the experimental results given in the last tables. The limiting value of $\frac{n}{N_1}$ is 1, while the value of $\frac{n}{N_1 - n}$ rises to infinity. From van't Hoff's conception of the nature of osmotic pressure and its quantitative relations to molar concentration of the solute, it followed, as he showed by faultless thermodynamical logic, that relative depression of vapour pressure of the solvent is exactly proportional to osmotic pressure in dilute solutions. He knew Raoult's results as to depression of vapour pressure, just as he knew the corresponding results as to depression of freezing-point; but not seeing how otherwise to reconcile these results with his own unfortunate theory of osmotic pressure, he assumed that the divergence between his theory and Raoult's results was due to the interfering effects of attraction between the molecules of solute when the concentration of solute became more than very small. It is evident that when the value of n is very small, and the solute molecules are not very large, the values of $\frac{n}{N_1}$, $\frac{n}{N_1 - n}$ and $\frac{n}{N}$ become almost identical. Hence, for very dilute solutions van't Hoff's equations seemed to agree with the experimentally ascertained facts. The results of his misunderstanding have been most unfortunate in their influence on the development of physical chemistry; and, largely owing to the brilliance of van't Hoff's logic, the experimental results of Raoult have fallen into the background. We have only to substitute $\frac{n}{N_1}$ for $\frac{n}{N}$ in van't Hoff's equation $\frac{p_d}{p} = \frac{n}{N}$, and to correct his confusion of osmotic depression with osmotic pressure, in order to reap the real fruits of his thermodynamical reasoning.

Before considering this reasoning more fully in connection with the boiling-points of solutions, it will be convenient to discuss the very simple argument devised by Arrhenius[1] to show that for dilute solutions $\frac{p_d}{p} = \frac{n}{N}$, so that p_d varies as n. Arrhenius imagines a tube standing vertically and ending below in the cell of an osmometer dipped in pure water or other pure solvent. The cell is filled with a very dilute solution of sugar or other solute, and the liquid also fills the tube to such a height above the level of the water that the extra pressure in the cell is exactly the osmotic pressure, so that water neither passes in nor out through the walls of the cell. The temperature is also assumed to be the same in the cell, tube, and surrounding air, and no air-currents are present; and the air is either saturated with the aqueous vapour given off from the pure water below, or no air is present.

---

[1] Arrhenius, *Zeitschr. f. physikal. Chemie*, iii., p. 115, 1889.

It is evident that for small differences between the level of the water into which the cell dips and the level of the solution in the tube, the differences between the vapour pressures at the two levels will be sensibly proportional to the difference in height between the two levels, just as in the case of differences of ordinary barometric pressure with small differences of altitude. The vapour pressure will, therefore, diminish in proportion to the height. On the other hand, the osmotic pressure will also vary sensibly as the height of the column, so long as the solution is sufficiently dilute. Arrhenius argued that the osmotic pressure must vary directly as the vapour pressure, and the vapour pressure at the top of the tube must therefore be in equilibrium with the vapour pressure in the air. If the result was otherwise, we should have a system in which water molecules were passing in one direction or the other through the system consisting of solution, pure water, and air. This movement could be made the source of a small but unending supply of external work if the temperature were kept constant. But according to the second law of thermodynamics this is impossible. Hence for sufficiently dilute solutions the vapour pressure at the top of the column must diminish in proportion to the height of the column above the surface of the pure solvent, and therefore in proportion to the osmotic pressure.

It is of course quite clear that the reasoning cannot apply to strong solutions, where the specific gravity of the liquid in the column may increase considerably, so that the osmotic pressure increases faster than the height of the column, and where, just as in the case of barometric pressure, the pressure of any vapour present in the air, on account of the great height of the column, diminishes more slowly than in direct proportion to the height. Nevertheless, in the case of weak solutions the reasoning is generally treated as valid.

There is a fallacy in the reasoning, and we shall see that when this fallacy is corrected, the reasoning leads to Raoult's, and not van't Hoff's conclusion. At constant temperature the vapour pressure in equilibrium with water *in vacuo* diminishes with height above the water in proportion to the diminution in the weight of vapour in the column of aqueous vapour of a given sectional area existing above the height at which the vapour exists. At a certain height (about 30,000 feet) the vapour pressure is diminished to half; at double this height to a fourth, and so on indefinitely. The weight of a given small length of the column depends upon the vapour pressure of water at the given temperature and the molecular weight of water. Thus the relative rates at which different vapours or gases diminish in pressure with height depend, not simply on height but upon height multiplied by their respective molecular weights; and this is so whether air is also present or not. But, as Perrin showed experimentally in the case of suspended particles, and as follows from calculations based on the extension to liquids of the dynamical theory of gases, the different constituents of a solution in a column of liquid obey the same law, and must do so whether the solution is dilute or concentrated. A homogeneous column of cane-sugar solution, for instance, is not in stable equilibrium, but if left to itself will become more concentrated at the bottom than at the top, because the molecular weight of sucrose pentahydrate is far greater than that of water. If the solution in the tube were homogeneous, there would as a matter of fact be no diffusion equilibrium at first between the sugar solution at the top of the column in the imaginary experiment of Arrhenius and the aqueous vapour in the air. Aqueous vapour would continue to pass over into the sugar solution at the top, and, as a consequence, water out of the sugar solution at the bottom, until equilibrium was established; or, if the sugar solution were at first left to itself for a sufficient time, the solution would become less concentrated to the required extent at the top than at the bottom to give equilibrium at once. The thermodynamical argument is therefore fallacious, since it makes the wrong assumption that the concentration of the solution in the tube is even.

Both the bombardment pressure of water in the solution and that of aqueous vapour in the air diminish with height of the column in the same proportion, determined by the molecular weight of water, and regardless of the presence of the solute or of air. At the bottom of the column the bombardment pressure of the water is P, and at the top $P - P_d$; and at the bottom the vapour pressure is p, and at the top $p - p_d$. Thus $\frac{P - P_d}{P} = \frac{p - p_d}{p}$ so that $\frac{P_d}{P} = \frac{p_d}{p}$. But P_d, as we have seen, equals $\frac{n}{N_1}$ P, and in the case of the top of the column this means the value of $\frac{n}{N_1}$ at that point. Hence $p_d = \frac{n}{N_1} p$ where the value of $\frac{n}{N_1}$ is that at the top of the column.

It is therefore to Raoult's and not van't Hoff's conclusion that the reasoning of Arrhenius, when fully followed out, leads. If, indeed, it were true that the vapour pressure varies simply as n, the second law of thermodynamics, or, as shown in Chapter II., the dynamical theory of heat, would be in gross contradiction with observation, and water would go on circulating for ever round the system and furnishing an endless supply of mechanical energy at the expense of heat in surroundings of the same temperature.

An interesting example of vapour pressure is that of a permanent gas dissolved in a liquid. When the gas dissolved in the liquid is in equilibrium with free gas in contact with the liquid, it is evident that the pressure of the gas is the vapour pressure of gas dissolved in the liquid. Now it has been found for various gases that the quantity of gas which goes into solution in a given mass of solvent is directly proportional to the pressure of the gas. This is Henry's well-known law of solution of gases. If we call n and N the respective molar concentrations of the gaseous solute and pure solvent, and p the pressure of the gas, then, according to Henry's law, p varies as n, and therefore as $\frac{n}{N}$, since N is a constant. By extending Henry's law to all cases of soluble vapours we might argue that $\frac{p_d}{p}$ must always be equal to $\frac{n}{N}$. From this we might further argue that osmotic pressure must be directly proportional to the molar concentration of the solute, as van't Hoff believed to be the case. This line of argument is, indeed, often used in support of van't Hoff's theory, and was so used by van't Hoff himself.

The truth is that all ordinary gases are, relatively speaking, so slightly soluble that it is impossible to distinguish experimentally between $\frac{n}{N}$ and $\frac{n}{N_1}$ as regards solubility. From analogy, however, we must conclude that the pressure of a gas on a solvent saturated with it varies within wide limits as the value of $\frac{n}{N_1}$ in the solution, just as Raoult found for the vapours of very soluble substances, where the values $\frac{n}{N}$ and $\frac{n}{N_1}$ diverged considerably. The more correct statement of the law is therefore that the molar concentration of the gas varies, within wide limits, as the ratio of moles of dissolved gas to moles of solution. This covers the case of gases which do not, at high pressures, obey Boyle's law.

Raoult investigated not only the depressions of vapour pressure of ether in solutions up to the moderate strengths for which data are given in the table on page 186, but also the depressions in very strong solutions. He found that, taking the results as a whole for each particular solute, the relative depression did not vary quite directly as the ratio of moles of solute to total moles in unit of solution, but roughly speaking as this ratio multiplied by a constant which varied for different solutes, and which was less than unity in all the cases investigated. The discussion of these results will be postponed to Chapter VII. Meanwhile we must pass on to elevations in the boiling-points of solutions.

Raoult discovered that, just as in the case of depressions of freezing-point, though not up to such high concentrations of solute, the elevations of boiling-point in solutions in various organic solvents vary with the ratio of moles of solute to a given weight (or given number of moles) of solvent. He took 100 grammes as his standard weight of solvent; but if we take 1000 grammes instead, it follows, as already pointed out (p. 112), that the rise of boiling-point is proportional to $\frac{nN}{(N_1-n)S}$, where S is the specific gravity of the pure solvent, and N its molar concentration, at its boiling-point.

The following table, compiled from the data collected in Landolt and Börnstein's *Tabellen*, illustrates the application of Raoult's law in the case of cane-sugar solutions. As Morse's observations on osmotic pressure indicated that cane sugar at 100° is not hydrated, the calculations are for unhydrated cane sugar.

Elevations of Boiling-Point of Cane-Sugar Solutions.

Grammes Anhydrous Sucrose per 100 g. Water.	Moles Sucrose per 1000 g. of Water $\left(\frac{nN}{N_1-n}\right)$.	Observed Elevation of Boiling-point. Degrees C.	Molecular Elevation (Observed Elevation divided by $\frac{nN}{N_1-n}$). Degrees C.
2·447	·0715	·035	·49
4·316	·1264	·064	·51
7·25	·212	·103	·49
11·02	·322	·164	·51
16·75	·490	·242	·495
21·68	·634	·317	·50
36·15	1·056	·55	·52
65·47	1·93	1·13	·59
103·4	3·02	2·06	·68
175·1	5·12	3·84	·75
276·2	8·07	6·71	·83

It will be seen that up to a concentration of about 20 grammes of sugar per 100 grammes of water, or about .6 mole per 1000 grammes of water, the elevation of boiling-point is proportional to the number of moles of sugar per 1000 grammes of solvent. At greater strengths the elevation is in increasingly higher proportion. This reminds us at once of the diminished rate of depression of vapour pressure at high concentrations of the solute; and the cause is presumably the same, namely, rise in the internal pressure, P, of the liquid. The explanation of this will be discussed in Chapter VII.

It is evident that, just as in the case of depression of freezing-point, the elevation of boiling-point is proportional, not to molar concentration as van't Hoff thought, but to moles of solute per unit weight, or unit number of moles, of solvent, as Raoult found by experiment.

If we assume, as in calculating depressions of freezing-point, that the internal bombardment pressure is the same in the solution as in the pure solvent, it follows that when n moles per litre of a solute without sensible vapour pressure are present, the reduction in diffusion outwards of the solvent vapour is $\frac{n}{N}(1 - Nb)p$ (see p. 193). To bring the solution to the boiling-point the vapour pressure of the solvent must therefore be raised by the amount (p_e) required to neutralise the deficiency and so raise the vapour pressure to the boiling pressure p. Hence $\frac{N_1 - n}{N}(1 - Nb)(p + p_e)$ must equal $\frac{N_1}{N}(1 - Nb)p$. It follows that $p_e = \frac{n}{N_1 - n}p$, or $\frac{p_e}{p} = \frac{n}{N_1 - n}$; or, since N is a constant, $\frac{nN}{N_1 - n}$ is to N as p_e is to p; or, since S (the specific gravity of the solvent) is also for small differences of temperature practically a constant, $\frac{nN}{(N_1 - n)S}$ is to $\frac{N}{S}$ as p_e is to p. Thus if p is the boiling pressure (one atmosphere) and there are n moles of solute to $N_1 - n$ moles of solvent in the solution, the vapour pressure must be raised to the extent of $\frac{n}{N_1 - n}p$ in order to bring the solution to the boiling-point. If, therefore, there were one mole or any other definite fraction or multiple of a mole of solute to 100 grammes of a given solvent, the vapour pressure would need to be raised to the same amount whatever the chemical nature of the solute: and since, at the boiling-point, p is the same for all solvents, we can extend the rule that $p_e = \frac{n}{N_1 - n}$ of an atmosphere to all solvents, as well as all solutes.

This does not tell us anything about the actual rise of temperature required to raise the solution to the boiling-point, since increase of temperature bears no simple relation to increase of vapour pressure. For a given solvent there is, however, an empirically-known relation between temperature and vapour pressure. Hence we can conclude that for any given solvent there must be a definite relation between ratio of moles of solute to moles of solvent and the elevation of boiling-point. That this relation actually exists has already been seen, and constitutes Raoult's law as to elevation of boiling-point.

It may be noted here that the significance of Raoult's algebraical notation as applied to molar concentrations of solutions is entirely different from that employed in this book. In Raoult's notation n and N mean in reality nothing more than numbers of moles of solute and solvent. They do not mean, as in this book, or in van't Hoff's thermodynamical calculations, numbers of moles, or molar concentrations per litre. Hence in Raoult's notation the ratio $\frac{n}{N}$ means the same thing as the ratio $\frac{n}{N_1 - n}$ in the notation of this book, and signifies ratio of moles of solute to moles of solvent. This ratio is the same in a litre or 100 grammes, or any other volume or mass which may be taken as a standard. Raoult's standard was 100 grammes of the solvent present, so that with his notation $\frac{n}{N}$ 100 meant moles of solute to moles of solvent in 100 grammes of solvent, while $\frac{n}{N+n}$ 100 meant moles of solute to moles of solution per 100 grammes of solvent. With the notation of this book the corresponding expressions would be $\frac{n}{N_1 - n}$ 100 and $\frac{n}{N_1}$ 100, or $10 \times \frac{nN}{(N_1 - n)S}$ and $10 \times \frac{nN}{N_1 S}$, where S = specific gravity referred to that of water at 4° as unity.

Why does the vapour pressure of a liquid solvent increase with rise of temperature? The reason is not that the internal bombardment pressure of the liquid increases; for we shall see later that it diminishes, whereas the vapour pressure increases very rapidly with temperature. For instance, the vapour pressure of water is 165 times greater at 100° than at 0°. The reason for increase of vapour pressure with temperature is that as the temperature rises, and with it the mean energy of translation possessed by its molecules, a larger and larger proportion of these molecules acquire the translational energy required to enable them to perform the work of freeing themselves from the surface of the liquid. With the eye of genius, van't Hoff saw that since the latent heat of evaporation per gramme of solvent at any temperature is a measure of the energy required to convert one gramme of solvent into its vapour at the same temperature, and since by the kinetic theory rise of temperature is a measure of the translational energy of the molecules, the proportion of solvent molecules set free in unit time must vary directly as T, and inversely as L, where L is the latent heat of evaporation in gramme-calories per gramme of solvent at the temperature T. Following this clue he calculated the constants which, as Raoult discovered, connect, in the case of each solvent, the elevation of boiling-point with the moles of solute present.

In making his calculation, van't Hoff assumed that osmotic pressure varies as the molar concentration of the solute, and made use of the machinery of an imaginary Carnot cycle and the so-called second law of thermodynamics. The calculation gave him approximately the constants which Raoult had discovered; and this has been regarded as a great triumph for van't Hoff's theory of osmotic pressure. Concealed in the machinery, however, there were two errors, one of which was very serious, but fortunate in the sense that it brought his calculated figures into much better agreement with the experimentally determined figures. The other was of less importance.

The calculation may be put into a simpler form, dependent directly on the dynamical theory of gases and liquids.

On the dynamical theory, as we have already seen, the translational energy of one mole of solvent at T is, when expressed in heat-units, 1.983 T calories. The latent heat of evaporation of one gramme of solvent is L calories at the temperature T. Hence the latent heat per 1000 grammes of solvent is 1000 L, and the translational heat-energy $\frac{\text{N}\ 1.983\ \text{T}}{\text{S}}$. The probability of a molecule of solvent flying off as vapour must therefore depend (within small intervals of temperature, within which L is practically constant) on the quotient of the second of these two quantities divided by the first, *i.e.* on $\frac{\text{N}\ 1.983\ \text{T}}{1000\ \text{LS}}$.

Now we have already seen that p_e is to p as $\frac{n}{\text{N}_1 - n}$ to 1. Hence when $p = 1$ atmosphere, as it is at the boiling-point, p_e varies as $\frac{n}{\text{N}_1 - n}p$. It is evident, moreover, that since the value of $\frac{\text{N}\ 1.983\ \text{T}}{1000\ \text{LS}}$ varies directly (over small intervals of temperature) as T, $\frac{p_e}{p}$ must vary as $\frac{\text{T}_e}{\text{T}}$ where T is the boiling-point of the pure solvent and T_e is the elevation of boiling-point per mole of solvent when a solute is present. Hence $\frac{\text{T}_e}{\text{T}}$ varies as $\frac{n}{\text{N}_1 - n}$. But it must also, as just seen, vary as $\frac{\text{N}\ 1.983\ \text{T}}{\text{S}}$. Hence T_e varies as $\frac{n\text{N}}{\text{N}_1 - n} \times \frac{1.983\ \text{T}^2}{1000\ \text{LS}}$, or $\frac{nN}{(N_1-n)S} \times \frac{1.983\,T^2}{1000\,L}$. But $\frac{nN}{(N_1-n)S}$ means moles of solute per 1000 grammes of water multiplied by N, and a litre of water at 100° contains only .9854 as many moles as 1000 grammes, or a litre of water at 4°. Hence when one mole of solute is present per 1000 grammes of solvent we must multiply $\frac{1.983T^2}{1000\,L}$ by .9584 in order to get the true molar elevation, which is simply the value of $\frac{1.983T^2}{100L}$ when the true value of $\frac{nN}{(N_1-n)S} = 1$. Hence the actual elevation T_e is equal to Raoult's molar elevation multiplied by S, or $\frac{1.983\,T^2S}{1000\,L}$.

For this value we can now substitute the actual calculated value for an aqueous solution. The value of S for water at 100° is .9584; T = 373.1°; and L = 536. Hence $T_e = \frac{1.983 \times 371.1^2 \times .9584}{1000 \times 536} = .495$.

When this calculated value is compared with the values found experimentally, as given for cane-sugar solutions in the table on p. 204, it will be seen that the agreement is very close for strengths up to at least 20 grammes per 100 of water. The third of the determinations (.49) is by Beckmann, who, as is well known, introduced many important improvements into the methods of determining elevation of boiling-point and lowering of melting-point. For a mannite solution he obtained .50; and for solutions of boracic acid, which are almost unionized, .49 to .50. For highly ionized substances, such as potassium or sodium chloride or hydrate, the molar elevations of very dilute solutions are increased in proportion to the number of ions formed, just as could be predicted from the conclusions of Arrhenius. Thus in the case of potassium chloride or hydrate the elevations are doubled, since for each molecule two ions are formed.

According to van't Hoff's calculation the molar elevation is .516. The difference between this and .495 arises from the fact that he took no account of the expansion of water by heating. The expression which he deduced for molar elevation (as defined on Raoult's system) was $\frac{n}{1000} \times \frac{1.983\,T^2}{L}$. This would be correct if the specific gravity of water were the same at 100° as at 4°, but actually introduces an error of about 4 per cent., which is the smaller of the two errors already referred to in van't Hoff's calculation. Morse, as we have seen (p. 157), made the same error in calculating the theoretical osmotic pressures of hot sugar solutions.

By a similar process of reasoning, van't Hoff predicted the molar depression of freezing-point. Here, however, the error of 4 per cent. did not exist, since to S is given its actual value at freezing-point. Thus, as already pointed out, van't Hoff's expression gives, with the utmost accuracy, the actual depression of freezing-point when *n* is taken in Raoult's sense, and allowance is made for the hydration of the sugar molecules (see table, p. 178).

We must now consider the other and far more serious, though in one sense fortunate, error in van't Hoff's calculation. According to his theory the osmotic pressure, and consequently the elevation of boiling-point and depression of freezing-point, vary as the molar concentration of the solute per litre or per 100 c.c. of the solution. In calculating osmotic pressure from the gas laws, he designated this concentration as *n*. On reading his thermodynamical deduction of the expression for Raoult's molar elevation or depression of temperature, it will easily be seen that, at any rate in his paper of 1887, he confused the *n* of his own notation with the *n* of Raoult's notation. Moles per 100 c.c. of solution and moles per 100 grammes (or 100 c.c.) of solvent present are two quite different things. Van't Hoff's *n* meant moles per 100 c.c. of solution, and Raoult's *n* meant moles per 100 grammes of solvent. Only at enormous molar dilution, and in a solvent of the specific gravity of water at near 4° C., do they mean practically speaking the same thing. To take an instance, the molar elevation given by a solution containing 10.18 grammes of (anhydrous) cane sugar per 100 grammes of water gives a molar concentration of 0.297 per 1000 grammes of water, but only of 0.281 moles per litre of solution. The former figure gives exactly the molar elevation (.515°) corresponding to van't Hoff's calculated expression; but a correct calculation on his own principles would have given .545°, instead of .515°

as the molar elevation. He did not realise that, even when he was referring to dilute solutions, he had taken Raoult's n for his own.

Van't Hoff, with all his great genius and knowledge of mathematics, does not seem to have been a very good calculator. Raoult left the calculation to Nature, and she did the sums in accordance with the dynamical theory. Van't Hoff calculated them on his own theory for himself—and the results of the sums, in so far as they were correctly done, did not agree with observation. But a lucky mistake gave him the right calculated result for depressions of melting-point, and nearly the right result for elevations of boiling-point.

It may be pointed out that at very great dilution it is impossible to distinguish by experiment between n, $\frac{nN}{N_1}$, and $\frac{nN}{N_1 - n}$. Hence at very great dilutions it is practically true that osmotic pressure, etc., vary with the molar concentration of the solute. Both van't Hoff himself and those who defend his equations have taken this line of argument, and have sometimes based a deduction of these equations on Henry's law, which says that the molar concentration of a dissolved gas is proportional to the pressure of the gas. In this case the molar concentration in question is so minute that for ordinary gases at ordinary pressures Henry's law is practically true. A further step in the argument is that as the concentration increases the solute molecules come so close together that, as was supposed by van der Waals to happen in the case of gases, the mutual attraction of solute molecules begins to blur the results. For this reason it was assumed that van't Hoff's equations no longer hold good exactly.

It may be granted that at sufficient dilutions we could not disprove van't Hoff's equations by experiment; but neither could we verify them as against Raoult's. On the other hand, they do not hold with any concentrations at which they can be really tested, while Raoult's conclusions, and the corresponding ones relating to osmotic pressure, do hold, and are inconsistent with van't Hoff's equations. There are thus no experimental grounds for a defence of the latter equations, however useful they have been as a temporary scaffolding in building up the science of physical chemistry as applied to solutions. That van't Hoff's equations are completely unintelligible physically has also been pointed out already.

We can see what misled van't Hoff into thinking, from Pfeffer's imperfect results with sugar solutions, that osmotic pressure varies in proportion to the molar concentration of the solute. We can also see how he failed to see that Raoult's results for depression of melting-point, elevation of boiling-point, and depression of vapour pressure, in solutions were inconsistent with his theory. It seemed to him impossible to obtain any work through the interaction of two chemically and electrically indifferent substances at equal pressure and temperature. He therefore concluded that the osmotic work which can be done when a solvent and a solution in it at equal temperatures are in contact through a semi-permeable membrane must be due to some added pressure; and since this pressure seemed to be proportional to the molar concentration of the solute and to the absolute temperature, as in a gas, he concluded that in some quite unintelligible manner there exists in a solution a pressure acting inwards, which varies in accordance with the equation $PV = nRT$, where P is osmotic pressure, and V is the volume of solution. He had to locate the pressure in the solution, and not in the solvent, because unless he located it in the solution he could see no way of making the equation correspond with the observations. The apparent fact of the correspondence was the big stick with which he could belabour the heads of his opponents, though how a positive pressure in a solution could drive water into it was as unintelligible as the most unintelligible of theological dogmas.

In spite of all his extraordinary genius, van't Hoff often applied mathematical expressions blindly. He ignored the significance of the b in Hirn's and van der Waals' equations. He also ignored the physical meaning of a positive pressure. The result has been confusion which it will still take years to recover from.

CHAPTER VI

THE LIQUID AND GASEOUS STATES

Enormous internal bombardment pressure in a liquid, without external pressure except for trifling vapour pressure, can, apparently, only be possible through the existence of powerful intermolecular cohesive "force," which prevents the molecules in the interior of the liquid from flying apart to more than a very limited mean distance from one another, though they are kept at this mean distance by the constant movement which corresponds to their temperature. Since this pressure is unbalanced externally they must behave as if they were compressed into a limited space by an elastic pressure externally applied, but were free to move about and jostle one another inside this space, owing to their mutual attractions being balanced.

Why, as van der Waals discovered, does the manifestation of the cohesive force vary, within wide limits, approximately in inverse proportion to the square of the gross volume occupied by a gas? Why do gases condense under pressure into liquids, and why is the condensation pressure so strikingly dependent on temperature? Why is there a critical temperature above which, as Andrews showed, gases cannot be liquefied? Why, as indicated in Chapters IV. and V., does the internal pressure remain the same in a liquid and its solutions up to considerable concentrations of the solutions? These questions, with others closely related to them, must now be discussed.

As already pointed out in Chapter I., the volume of a molecule can only be a virtual or statistical volume, by assuming which we can for practical purposes predict the manner in which immense numbers of molecules, when considered collectively, behave towards one another. We have no reason to believe that even an electron is "solid" in the old sense; and the old conceptions of "occupied" and "empty" space, or of solid and porous bodies, have melted away and become as shadowy as Kant's "noumenal" world. Those conceptions, however useful they may be in ordinary practical life, cannot represent reality. There remains for the present, however, the conception of the "mass" of molecules and atoms, together with that of attractive or repulsive "forces" associated with this mass, and the conception of "energy."

In connection with molecular physics, it now becomes necessary to frame some hypothesis as to how the relations between the attractive and repulsive forces involved alter with the distance of molecules from one another. The hypothesis which has been forced on me by the facts and conclusions discussed in the two preceding chapters, and by many other facts, will now be stated. As will appear below, this hypothesis is at any rate a peg on which a great mass of facts can be hung.

In each normal molecule or atom a very intense attractive "force" acts on other molecules or atoms from a relatively central position, and a *slightly less* intense repulsive force acts from what may provisionally (until the solid state is considered) be called a sphere outside the central position, and representing the location in the atom and on its surface, of "negative" electrons. Each of these forces varies, at a sufficiently great distance from the central point, inversely as the square of the distance from this point. As a consequence of the repulsive force originating from an external shell, this force, though smaller on the whole, must increase more rapidly than the attractive force when two molecules approach very close to one another, and will therefore at a certain distance neutralise, or cut across, the attractive force, thus defining a virtual surface (not the statistical virtual surface in a dense collection of molecules, since in such a collection the bombardment pressure effects the virtual surface) at which attraction gives place to repulsion, or within which kinetic begins to be transformed into potential translational energy as two molecules approach one another.

Figure 7, representing the manner in which the two supposed forces may be assumed to vary as the distance in one plane from the centre of a molecule C, will serve to illustrate this conception. It will be seen that as two molecules approach one another the net attractive force first increases as the square of their distance apart diminishes, then gradually ceases to increase, and then diminishes rapidly to zero. With still nearer approach there is rapidly increasing repulsion. It is evident also that the space between the two curves from any distance indicated by an ordinate may be taken, when two molecules approach one another, to represent net energy of attraction on one side of the zero-point, and net energy of repulsion on the other.

When the two molecules are at such a distance from one another that the repulsive sphere acts just as a point would, the residual attraction will simply vary inversely as the square of the distance, just as gravitation does. We can thus attribute gravitation to this residual net attraction, which will only be appreciable at ordinary sensible distances when the joint attraction of enormous collections of molecules is concerned. On the other hand, what we call electrical forces can be attributed to abnormally

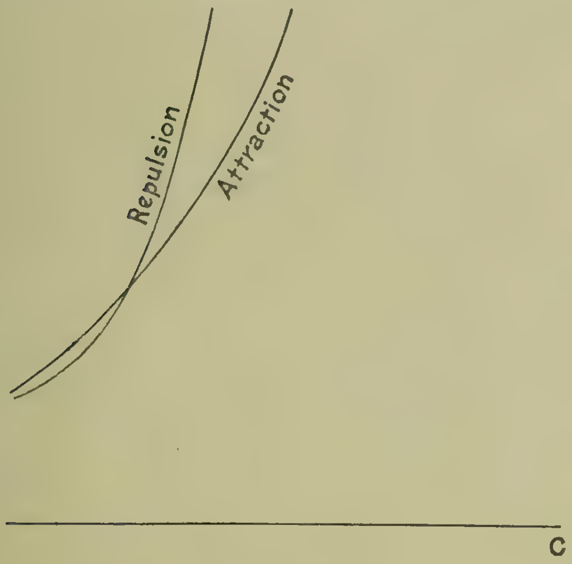

Fig. 7.

unbalanced attractions and repulsions. It is only when electrons become displaced owing to molecules being forced into one another or torn in pieces that we have any opportunity of realising the tremendous intensity of the forces of which gravitational and molecular attractions are only a residual manifestation, and chemical attraction a partial manifestation. In proportion to these unbalanced forces the normal residual difference between them seems so trifling that we ordinarily regard positive and negative charges as completely neutralising one another. The product of their union is "electrically" neutral; but from a wider point of view this product is by no means neutral to surrounding matter.

After writing the first draft of this chapter, I discovered that a somewhat similar conception, originating from Mosotti, had been discussed by Waterston in 1843, two years before he sent his famous paper to the Royal Society. The passage will be found at p. 168 of his *Collected Papers*. It will be seen, however, that at the time when he wrote he still adhered to the old idea (associated with the caloric theory) that the molecules of a gas, as distinguished from a liquid or solid, repel one another owing to a preponderance of repulsive force between them.

It will be evident that the hypothesis just outlined does not imply any direct relation between the intensities of gravitational and molecular or atomic attraction. The distance of the outer shell from the centre, as well as the relative intensities, of the two forces shown in the diagram determines the intensity of molecular attraction, while only the differences in intensity determine gravitational attraction. Of course, also, the diagram tells us nothing as to the connection between gravitational and inertial mass, and implies no assumption as to what "force," "mass," and "energy" are. Its function is only to help in the prediction of molecular behaviour.

It is evident that when molecules start to approach one another from a distance which is within one another's spheres of sensible attraction, the kinetic energy which each molecule gains in approaching another will be diminished in proportion to the diminished amount of acceleration. Within a closely packed mass of molecules, however, the acceleration will be more or less neutralised by retardation due to the close propinquity of the molecule or molecules with which the last collision has occurred. With sufficient molecular concentration the retardation, as will appear from Figure 7, must become equal to the acceleration in the middle part of a molecular vibration. Within the mass the molecules will thus, so far as their mutual attractions are concerned, be free to move about to a considerable extent in various directions, but will tend to be held together by their mutual attractions.

The influence on internal bombardment pressure of variation in the gross volume occupied by a gas or other collection of molecules has now to be considered. When molecules collide with one another, the kinetic energy of translation or *vis viva* which they possess at any moment, and which, in accordance with the kinetic theory, determines the pressure which they produce per unit of cross-sectional area, depends partly on their original energy of translation, and partly on the accelerations which they have received in virtue of their mutual approach. The latter accelerations may be, relatively to their original velocities, small or great. In the case of a substance which, in spite of great molar concentration, retains the gaseous form at all easily attainable temperatures, the acceleration must be relatively small; whereas in a substance which cannot be gasified at any easily attainable temperature the acceleration must be very great. Helium and carbon might be instanced as extremes in the first and in the second directions. It is evident also that the amount of acceleration, which depends on the time during which it acts, will depend inversely on the original velocity of the molecules.

Now if we diminish the gross volume occupied by a given number of gas molecules we correspondingly increase, not only the concentration of molecules ready to collide with other molecules, but also the concentration of those ready to be collided with. Hence the frequency of collisions will depend on the square of the molecular or molar concentration. The extra kinetic energy, and therefore the extra pressure per unit of cross-sectional area, due to acceleration as the molecules approach one another must thus vary directly as the square of the molar concentration. This reasoning is of the same character as that of Guldberg and Waage in their explanation of the influence of concentration of the reacting substances in a solution on the velocity of reaction.

If, therefore, we call π the extra internal pressure produced at a given temperature for each mole of gas present per litre, *i.e.* per unit value of N, the extra internal pressure produced at the same temperature, assuming that the molecules approach one another from a distance beyond that at which their mutual attraction is sensible, will be, when N increases, πN^2. The equation for a gas which is varied in concentration but is kept at a constant temperature, becomes therefore, if p is the external pressure which it exercises, $(p + \pi N^2)(1 - Nb) = N\ .082\ T$.

It is evident from this equation that T, which, on the kinetic theory, is a measure of the bombardment energy of the gas molecules, if it were not kept constant by abstracting or adding heat, would rise or fall with $\pi N^2(1 - Nb)$, as well as with N. If, however, we prevent this rise or fall by removing or adding heat, which means diminishing or increasing the bombardment energy which the molecules would otherwise possess, the equation would give the external and internal pressures with T constant.

We thus see that P, the total internal bombardment pressure, is never the same as p, the pressure exerted externally. The former pressure varies directly as the molar concentration per litre N, and inversely as the virtual free space $1 - Nb$, if there is no gain or loss of molecular kinetic energy; whereas this is not so for the latter pressure, since inside the gas the frequency of collisions increases as the square of the molar concentration, whereas the collisions with the walls of the containing vessel only increase directly as the concentration. The difference is quite unaffected by whether the walls of the containing vessel attract the gas molecules or not.

We can also see that constancy of the value of π in the equation depends upon whether the extra translational energy acquired by a molecule in approaching another molecule remains constant with changes of molar concentration. This can only be the case if the distance from which the molecules approach one another is outside the range of their sensible mutual attractions. If we regarded the molecules as being already within the range of their sensible mutual attractions the constancy could not hold, since the translational kinetic energy which they acquired in virtue of acceleration would diminish as their mean distance apart diminished.

If we use the older form of notation, in which only one mole of gas is assumed to be present, so that the fundamental equation takes the form $P(V-b)=RT$, the equation just arrived at takes the form $\left(p+\frac{\pi}{V^2}\right)(V-b)=RT$, since N varies inversely as V, so that $\frac{1}{V^2}$ corresponds to the N^2 of the new notation. If for π we substitute a, the equation becomes identical with van der Waals' equation $\left(p+\frac{a}{V^2}\right)(V-b)=RT$. We can thus deduce van der Waals' equation if, but only if, we assume that the molecules approach one another from outside their spheres of sensible mutual attraction, and that b is a constant.

When he discovered, through mathematical analysis of Andrews' experimental results, the facts which are embodied in his famous equation, van der Waals concluded that the reason why the external pressure is less than it would be if Boyle's law, in the form given to it by Dupré and Hirn (see p. 14) held good, is that an attractive force, *a*, which varies as the square of the concentration, and therefore inversely as the square of the volume, is coming into play, and thus diminishing the external pressure which would otherwise exist. He conceived the attractive force as varying inversely as the volume, and also directly as the concentration, which likewise means inversely as the volume, so that the total diminution of external pressure varies inversely as the square of the volume.

This hypothesis may seem plausible at first sight, but is totally inconsistent with the dynamical theory of gases. If van der Waals' theory were correct, the internal bombardment pressure of a gas, and therefore its volume at any given temperature, would simply vary in proportion to the external pressure, since the extra kinetic translational energy, or *vis viva*, acquired by the molecules in their approach would, as shown by the reasoning stated above, diminish in proportion to the square of the concentration, or in inverse proportion to the square of the volume. The assumption of van der Waals is thus quite impossible, as it leads to a conclusion contradicting the facts. As will be shown below, the theory of van der Waals treated gases as if they were already liquids, and it could thus give no account whatever of condensation to the liquid form, or of a critical temperature. Because the a of van der Waals' equation suggests, not pressure, but attractive force acting at a considerable distance, I have not retained his lettering.

It seems extraordinary that though van't Hoff's interpretation of his discovery of the approximate agreement for dilute solutions between the laws of osmotic pressure and Boyle's, Avogadro's, and Charles's laws has constantly been questioned, there has been general acceptance of van der Waals' interpretation of the equation which he discovered from Andrews' experimental results. Doubt seems to have been stifled by the remarkable practical success of the equation. It was also a fascinating mathematical plaything. It is, however, from a dynamical, and not a mere mathematical, point of view that we must approach the phenomena of compression of gases, as well as of osmotic pressure. From the dynamical standpoint the explanation of the squared volume term in the equation for a real gas can only be that given above. The assumption that the attractive force between molecules varies inversely as the square of the volume, would make a dynamical interpretation of the facts quite impossible.

Although the equation $(p+\pi N^2)(1-Nb)=N.082T$ is a simple one, we cannot assume that even at constant temperature either b or π is a constant, though at low pressures these assumptions are doubtless practically true. Since $p+\pi N^2$ (or P) is constantly increasing with increase in the value of N, the statistical value of b must, on the theory developed at page 17, Chapter I., be constantly diminishing, owing to the increased pressure. Some of the molecules, moreover, must approach one another from less distances since the last collision than would correspond to the full value of π; and this must occur to a constantly increasing extent as a gas increases in concentration, so that the value of π must be regarded as becoming increasingly smaller. Thus the value of π becomes increasingly smaller than it would otherwise be, while the value of $1 - b$ becomes increasingly larger. To a certain extent these changes cancel one another out; and this makes the equation a far more valuable one for practical prediction than it would otherwise be.

In all actual gases, as we have seen (page 223), heat becomes to some extent "latent" as a gas expands without doing external work.[1] The physical interpretation of this is, of course, that kinetic heat-energy is converted into potential energy of position. But according to the theory developed above, the molecules of a gas at any moderate pressure are on an average far outside of the distance from one another at which their mutual attractions would be sensible, whereas the heat which becomes latent as a gas expands without doing external work may be very great, as in the case of concentrated steam, ammonia, or carbon dioxide. The latter consideration probably helped greatly the general acceptance of van der Waals' theory of his equation. We can readily understand why heat should become latent as a gas expands without doing external work if we assume that in the gas the molecules are so close as to be on an average within their spheres of sensible mutual attraction. It is not at once evident that this assumption is wholly unnecessary.

---

[1] That hydrogen heats itself slightly in expanding into an almost vacuous space appears to be due to the imperfection of the vacuum.

The assumption is based on a very common error in statistical reasoning. It is true that on an *average* the molecules of a gas, except under very great pressure, are outside one another's spheres of sensible attraction; but in the existing chaos of molecular movement some of them (and the number of these per mole of any particular gas must vary inversely as the square of the volume) are nevertheless at any moment within the sphere of attraction, and in virtue of this fact must possess more kinetic and less potential energy. Kinetic energy is converted into potential energy in proportion as the number of those in this position is diminished owing to expansion. As regards this conversion it is practically of no importance how far, relatively speaking, the immense majority of the molecules are from one another. It is only those which are for the moment within spheres of mutual attraction that count in producing the effect; and they count very heavily on the total value of P if, as in the case, say, of ammonia, the value of π is relatively high.

The fallacy is of the same nature as that which, in physiology, has led to the conclusion that, assuming equilibrium to occur between the gas pressures in the lung alveoli and the blood leaving them, the gas pressures in the mixed arterial blood will be the same as the average gas pressures in the lung alveoli. Both of these values can be measured accurately, and they have been found not to agree, and sometimes to disagree widely. But if, as many physiologists at present do, we assume their agreement, we may be led to conclusions which are quite misleading—for example, to contradiction of the direct evidence that at high altitudes and under various other physiological conditions there is active secretion of oxygen inwards into the blood from the alveolar cavities.

Let us now consider what, on the dynamical theory as so far formulated, the distribution of externally disposable translational energy must be among the individual molecules of a gas. When two molecules, starting without relative movement in the direction of the axis uniting them, move towards one another under the influence of mutual attraction, or afterwards move away, they have no externally disposable translational kinetic energy in virtue of their mutual movement in the axis uniting them. Not only is this so, however, but to the undisposable kinetic energy which they possess in virtue of their mutual movement in the axis uniting them a corresponding amount is added of any kinetic energy which they possess in virtue of their mutual movements in the other two axes by reference to which their mutual movements can be expressed. This added amount becomes rotational energy, and is thus unavailable externally for producing pressure. Molecules have three degrees of translational freedom, and kinetic energy is bound in respect of them all as molecules acquire kinetic energy in virtue of their approach to one another. It follows that the more frequent the opportunities of molecular approach the greater the proportion of the total translational energy that becomes unavailable for producing external pressure—in other words, the less the external pressure that this energy can produce.

The reduced rate of rise in external pressure when the volume of a gas is diminished and its temperature is kept constant, is thus a measure of the degree to which the externally available translational energy is diminished by molecular attraction. It does not matter whether we say that the extent to which the value of p is limited by πN^2 is due to the increasing influence of molecular attraction or to the increasing proportion of molecules which at any given moment are held within one another's spheres of mutual attraction and can, in respect of the kinetic energy which, owing to this position, is unavailable externally, produce no external pressure.

If the quantities of kinetic energy corresponding to each degree of freedom varied together in an orderly manner in a gas, it is evident that there would be no external pressure unless the total translational energy of the molecules was greater than that held in leash by the orderliness of their mutual movements. In a well-executed dance there is no pressure against the walls of the room, however great the total translational energy of the dancers. But the dynamical theory, since it starts from the conception of molecular chaos, assumes that a gas is nothing but a quite disorderly assembly of moving molecules, in which the energy corresponding to each degree of freedom is, statistically regarded, distributed in accordance with mere abstract probability, and is indefinitely divisible. In respect of each degree of freedom some molecules will have more and some less translational energy than can be bound by molecular attraction. On this assumption the gas will still have a definitely calculable external pressure, however high the value of πN^2 may be; and when πN^2 equals p, equal quantities of translational energy will be bound by molecular attraction and free to exercise external pressure.

We have hitherto been considering the gas equation of state for the particular case in which the temperature remains constant, while the concentration of the gas is changed. It is now necessary to consider the case in which the temperature also varies. So long as a substance is in the gaseous state its molecules are constantly shooting quite clear of one another. The velocity with which they impinge on one another is thus composed, partly of the original velocity which they possessed when they were flying clear, and partly of the additional velocity which they acquire owing to acceleration during the period after they come within the spheres of their mutual sensible attractions.

The time during which the acceleration acts, and consequently the increased velocity itself, will be less the higher the temperature; but, on the other hand, the number of impacts will be correspondingly greater. Hence temperature by itself cannot alter the value of π. The total internal pressure, however, varies, at constant volume, directly as the temperature, and as this affects the virtual volume of the molecules, the value of b must diminish with temperature, so that the value $1 - Nb$ increases with increase in the temperature. If, therefore, we assumed, as van der Waals did, that b is constant, we should also have to assume that π (or the a of van der Waals' equation) diminishes with temperature; and various modifications in this direction have been made to van der Waals' equation. One simple modification of this kind is the equation of D. Berthelot, which, however, is not very successful. The truth, however, is that the value of b in the equation must be regarded as varying, however slightly, with pressure, just as the value of both π and b must, for the reasons already stated, be regarded as varying with molecular concentration.

Since at low molecular concentrations of a gas the value of Nb is extremely small, so that variations of the value of b with temperature make hardly any difference in the values given by the equation, we can neglect these variations at low molecular concentrations. We can, indeed, neglect the value of Nb altogether at low molecular concentrations, since its presence in the equation makes very little difference. The equation then becomes $p + \pi N^2 = NRT$. The corresponding form of van der Waals' equation becomes $pV + \frac{a}{V} = RT$, where V is the volume occupied by one mole. As the concentration increases, however, the value of Nb, and the diminution in the value of b with pressure, become of greater and greater importance, even though the value of π is so great that πN^2 cannot be neglected at low concentrations.

Owing to the fact that not all the molecules of a given mass of gas approach one another from starting-points outside the limits of sensible attraction, and that the number of those which do so diminishes with increase in the concentration, while b diminishes as the pressure increases, the values of π or a, and of b, in the equation of state, cannot be constants. If, through not understanding that this arises from their physical meaning, we expect them to yield an equation which will enable us to predict the results of unlimited changes in temperature and pressure, the only possible result is disappointment.

We have seen that, owing to the influence of diminution in the value of a being to a large extent cancelled out by diminution in the value of b, van der Waals' equation, or the corresponding one $(p+\pi N^2)(1-Nb)=NRT$, when applied to a gas at constant temperature but varying pressure, is particularly valuable for purposes of practical prediction over very wide limits of pressure. In the case, however, of variation of temperature with volume constant π does not vary with b. Hence when a concentrated gas is heated the fact that b is not a constant becomes manifest much sooner than is the case when T is constant and p varies.

The equation $(p+\pi N^2)(1-Nb)=N\ .082\,T$, when $N\ .082\,T$ is expressed in calories as $NT \times 1.983$ (p and π being of course expressed in atmospheres) enables us to calculate, assuming that π and b remain constant, the total heat in calories which disappears when a litre of gas expands against atmospheric pressure. From the relative values of p and πN^2 we can calculate the proportion of this heat loss due to internal work on the gas. It is evident also that from the observed fall in temperature when the gas is allowed to expand without doing external work, we can calculate the value of π, when the molar concentration N is not too high.

So far we have only partially considered what must happen when two molecules come within one another's spheres of sensible mutual attraction. It must, however, come about if, as assumed, a gas is a chaotic assemblage, that two or more molecules have, for the moment, insufficient mutual translational energy to enable them to separate, since while within one another's spheres of attraction they have lost part of their energy to another molecule or molecules. This will happen more and more frequently in proportion as N^2 increases or T diminishes.

Now, when several molecules are simultaneously held within one another's spheres of sensible attraction those possessing any excess over the average translational energy in respect of any particular degree of freedom can, by mutual impact, dispose of it to a greater or less extent to the others with a deficiency of kinetic energy. In this way the kinetic energy in the aggregated molecules tends to become averaged; and the larger an aggregate, the more effective will this averaging be. The aggregates tend, therefore, to become dynamically stable at the point where the average translatory energy of their molecules equals that of the surrounding free molecules. This limit of average kinetic energy must evidently depend on the number and kinetic energy of the free molecules and groups of aggregated molecules which are constantly colliding—in other words, on the value of p. But if the temperature is kept constant as the gas is compressed, more and more of the kinetic energy of the free molecules is removed, as already pointed out. Hence, as soon as the value of πN^2 in an aggregate equals p the aggregate will be stable, and will evidently become so with a lower value of N the higher the value of π.

Within an aggregate the value of N^2 is increased but that of π is diminished. An aggregate of sufficient size for averaging of its energy per molecule to occur will not only tend to become stable dynamically, but will absorb the smaller aggregates. As soon, moreover, as stable aggregates begin to be formed, p will cease to rise as N increases, since p cannot rise above the value of πN^2 in the aggregates. The larger aggregates will also swallow up unstable small ones. In other words, the gas will separate into liquid and vapour, and N can no longer increase in the vapour, so that as the volume is further diminished the condensation will continue till the whole of the vapour is condensed. We can thus understand the nature of condensation and the general effects upon it of pressure.

If, with continuous increase of N, πN^2 can never catch up p in the equation $(p + \pi N^2)(1 - Nb) = N .082 T$, there will evidently be no condensation, since the aggregates can never become dynamically stable, will only exist momentarily, and will not absorb one another. If π were a constant, or anything like a constant, it follows from the equation that πN^2 must ultimately catch up p, however great the value of T. But we have seen already that in proportion as molecules come more and more to be within the spheres of their mutual attraction the value of π must diminish. At great molar dilutions this diminution is hardly appreciable, so that π can in this case be treated as practically a constant. But when molecules begin to approach one another closely π begins to diminish in value more and more rapidly, since, for the reasons already given, the kinetic energy which molecules gain in approaching one another will fall off greatly, even though, *on an average*, individual molecules may be too far off one another to exercise sensible mutual attraction. At the same time, as we have already seen, b will also diminish with increasing rapidity as N increases. The diminution in π will tend more and more to check the limitation of the rise of p by rise of πN^2 to equality with it. By reference to Figure 7 it will be seen that as the molecules are forced together a point must come when with further approach π must diminish in proportion as N increases. Hence if T is sufficiently high to prevent πN^2 from catching up p before this critical stage is reached, no further diminution in volume will enable condensation to occur. Below a certain temperature, however, there will quite evidently be condensation, since any diminution of volume must, in accordance with the equation, increase πN^2 relatively to p.

We can thus predict a critical temperature above which no increase of pressure can produce condensation, and a critical pressure needed to produce condensation at just below the critical temperature. At the critical pressure and temperature the intrinsic pressure due to the kinetic energy acquired by the molecules in virtue of their attraction for one another is just equal to the external pressure; and if we know the critical molar concentration (or critical volume occupied by one mole), the critical temperature, and critical pressure, we can calculate from the equation (since $\pi N^2 = p$, and $\pi N^2(1 - Nb) = \frac{1}{2}N\ .082\ T$) the values of π and b at the critical temperature and pressure. These values, however, will not be their values at other pressures and temperatures. This is why no amount of blind manipulation of van der Waals' equation has ever led to a satisfactory result as regards the critical temperature. Abstract mathematical reasoning alone is quite insufficient in this case.

Stable aggregates must always be formed as soon as the volume is reduced to the point where, at the existing temperature, πN^2 becomes equal to p. The further the temperature is below the critical temperature the sooner, evidently, will this point be reached, since the value of πN^2 is greater relatively to that of p the lower the temperature. But this does not explain ordinary condensation, which occurs at a pressure where the value of πN^2 (or of $\frac{a}{V^2}$ in van der Waals' equation) is far below that of p. Let us see, however, what must happen when, by whatever means, a liquid aggregate of considerable size has been formed when the temperature is below the critical point.

In the aggregate itself the molecules are, relatively speaking, closely packed together, so that the value of N^2 is far higher than in the gas, and $1 - Nb$ much lower. On the other hand, the value of π, since the molecules of the aggregate are within their spheres of sensible mutual attraction, is less than in the gas. It will, however, be shown in the sequel that π cannot be diminished in greater proportion than N is increased, so that at anywhere below the critical temperature and pressure (above which dynamically stable aggregates cannot exist) the value of πN^2 in an aggregate is higher than that of p, and as the temperature is lowered without altering the pressure of the gas, becomes, relatively speaking, higher and higher than that of p, which diminishes far more rapidly than in proportion to $1 - Nb$. We can easily see this from the equation $(p + \pi N^2)(1 - Nb) = NRT$. In this equation πN^2 does not diminish at all as T diminishes, and $1 - Nb$ only diminishes as N increases owing to the fall in temperature. Hence p must diminish to a smaller and smaller fraction of πN^2, and far more rapidly than T, as T is progressively lowered. With perfectly free molecules, on the other hand, p would only diminish in proportion to T as T was lowered.

But p, the external pressure of the liquid aggregate, is simply its vapour pressure. Hence the vapour pressure of the liquid aggregate is less than the external pressure which the remaining gas would have at the same pressure and temperature if the liquid were not present. The necessary result is that the liquid absorbs gas molecules until equilibrium is established between gas pressure and the giving off of molecules from the liquid. The gas, therefore, becomes more tenuous, and at low temperatures far more tenuous, than it otherwise would be. In other words, the value of N in it is reduced, and the value of πN^2 becomes less, and may be far less, than the value of p. Thus at what we know as the ordinary condensation-point of a gas the value of πN^2, or of the $\frac{a}{V^2}$ of van der Waals' equation, may be far less than that of p. In aqueous vapour at ordinary atmospheric temperatures, for instance, the value of πN^2 in the equation for the vapour is so small as to be quite negligible, and even at boiling-point is very small. The conditions on which the variations in the value of N with temperature depend, will be discussed at the end of Chapter VII.

In the condensation of a gas or vapour, whether the condensation aggregates are small or large, heat is necessarily produced, or, to use the nomenclature of the caloric theory, "liberated." When a molecule becomes aggregated with another owing to the accident of a collision with a third molecule while the two molecules are within one another's spheres of attraction, the third molecule has taken up kinetic energy representing what the other two would have lost through conversion into potential energy if they had rebounded from one another normally. Thus the sum of kinetic energy or heat energy in the free gas is increased, though quickly lost again if the temperature is kept constant, or if unstable molecular aggregates are being as often broken up as formed. When a gas is progressively condensed owing to its external pressure being kept just above the vapour pressure of its liquid the heat is produced at the surface of the liquid, and represents molecular potential energy which has been turned into the molecular kinetic energy called heat. Unless its temperature is kept constant the liquid will of course heat up, and the rise of temperature affords a means of measuring the heat produced.

On the caloric theory this heat was "latent" in the steam, and only became "manifest" on "liberation" of the heat in condensation. In a condensing steam-engine the heat is "liberated" at the condenser surface, and not to any appreciable extent at the piston surface, whereas in the condensation or compression stroke of a gas-engine heat is "liberated" at the piston surface. In Chapter II. we saw how failure to realise this led to Carnot's mistake about his engine being reversible, and how this mistake has obscured the theory of heat-engines up to the present day.

It is well known that in the absence of either liquid or solid nuclei in a body of vapour the vapour can often be either cooled or compressed to a very large extent without any condensation occurring. It becomes "supersaturated," and no phenomenon in Nature is more interesting and significant than that of supersaturation. Supersaturation can be demonstrated for aqueous vapour in a striking manner with the help of the apparatus devised by the late Dr Aitken. This consists of a tube, lined inside with wet filter-paper, and with flat glass ends. The tube, which can be closed air-tight, is connected with a small air-pump, so that the pressure can be diminished, thus producing fall in temperature of the contained saturated air, with consequent condensation of moisture. If ordinary air is introduced through a tap into the tube, and the pressure then lowered, a cloud of condensed moisture fills the tube, so that the view through it is completely obscured. If this cloud is allowed to settle, and the experiment repeated without introducing unfiltered outside air, the cloud will be coarser, and will settle more quickly; and with further repetitions there will finally be no cloud at all, all the nuclei having been exhausted. The vapour simply condenses on the walls of the tube, and not in the mass of air, which becomes temporarily supersaturated to a very high degree. By ionization of air, as C. T. R. Wilson discovered, we can produce large numbers of condensation nuclei, the ions acting as nuclei. Aitken pointed out the unpleasant and dangerous consequences which would ensue if there were no condensation nuclei in the atmosphere.

It is because of the peculiar conditions existing in vapours in contact with their own liquids that we cannot calculate the condensation temperature from the equation of state for the vaporised substance. In the case of many solids and liquids, the molar concentration of the vapour becomes so low at ordinary temperatures as to be quite inappreciable by existing methods. We are so accustomed to this fact that we are apt to accept it without any thought as to its real significance.

Although no full discussion of the solid state will be undertaken in this book, it is necessary to make some reference to the passage of liquids into the solid state. In this chapter molecules have been treated as if their virtual forms might be spherical. If this were the case, we could not explain solidification. It is evident that at any moment the net attractive force at a given distance from the centre of a molecule differs in different directions, so that the molecule possesses poles of inferior or superior repulsion and virtual shapes corresponding to these poles. As a consequence of this they tend to fit on to one another and become united pole to complementary pole when the temperature is sufficiently low. The gross volume, N*b*, which they occupy when united is not, however, necessarily lessened, and may be increased, as when water or iron solidifies. The product of union is either a so-called molecular compound, or a crystal, or an irregularly formed solid.

Solid nuclei play exactly the same part here as liquid or solid nuclei in the condensation of gases, and supersaturated liquids can be produced in the same way as supersaturated vapours. The heat which becomes "manifest" when a liquid solidifies, or a molecular compound is formed, is due to the extra kinetic energy developed when molecules unite pole to pole without subsequently separating as is always occurring in the preceding liquid state. The forms taken by crystals are evidently due to the regular apposition of molecules pole to pole, whereas when the apposition is irregular and incomplete a non-crystalline solid is produced, and this may be an aggregate of several kinds of molecules. As a crystalline solid or a molecule of a molecular compound leaves no accessible poles of attraction within its substance, it is chemically pure apart from any foreign material which may be adsorbed on its free surface. In a finely porous substance like charcoal or coal, the free surface may be enormously extended.

If it were the case that solids take up and give off heat according to the same "laws" as gases and liquids commonly do, the scientific treatment of solids on Newtonian lines would present no greater difficulty than that of gases and liquids. The inadequacy of Newtonian conceptions as applied to solids becomes, however, so prominent that no attempt will be made in this book to deal further with the subject of solids.

CHAPTER VII

THE INTERNAL PRESSURES AND SATURATED VAPOURS OF LIQUIDS

WE are now in a position to discuss more fully the internal pressures of liquids. If the value of π in the equation of state $(p+\pi N^2)(1-Nb)=NRT$ were still constant in the state of concentration corresponding to the liquid state, a collection of molecules of any of the ordinary liquids at an ordinary temperature would be a relatively dense mass, with a vapour pressure which would be insensible, since the collection would absorb practically all the vapour in its neighbourhood. The theory formulated at the beginning of Chapter VI. (p. 217) gives us the reason why π cannot remain constant, and must rapidly diminish with increase of molar concentration in proportion as the molecules are, on an average, more and more within the range of sensible mutual attraction (apart from what we call gravitational attraction).

On this diminution in the value of π depends, as we have seen, the existence of critical temperatures and of all the characteristic properties of liquids. The manner in which π diminishes as molar concentration increases is of great importance, and the theory enables us to interpret this manner of diminution.

At the critical temperature, as we have seen, $p=\pi N^2$, so that $P = 2p$ or $2\pi N^2$. On an average the translational energy of the molecules is just sufficient to carry them clear of one another, so that the aggregates of which the gas is full are fugitive, and cannot grow at the expense of one another or of free molecules. With the smallest fall of temperature large aggregates become dynamically stable and absorb one another; and since at any moment half the molecules have insufficient translational energy to fly free from one another the collection separates itself out into almost exactly equal parts of liquid and gas, with almost exactly equal molar concentrations, though of course the liquid is very slightly more concentrated than the gas. The latter has lost half its molecules, but is also reduced to just over half its volume, and is still, therefore, full of fugitive aggregates, while the former has hardly any capacity for absorbing the gas, since the mean translational energy of its molecules is scarcely less than what would gasify them. With further fall in temperature the excess of stability in the liquid, and consequently its absorptive capacity, increases progressively, as already pointed out, with consequent rapid fall in what would otherwise be the concentration and pressure of the saturated vapour.

From Fig. 7 we can easily see that at first the fall of temperature below the critical point will increase very greatly the density of the liquid, since a considerable diminution in the mean amplitude of the movement of each molecule (or its mean free path) will be needed in order to correspond to a given diminution (equivalent, say, to one degree) in the mean translational energy which the molecules gain in colliding with one another. Hence the value of N in the liquid will at first increase very rapidly with fall in temperature, while that of π diminishes slowly. As a consequence the diminution in volume of the liquid must be very rapid at first. The great effect of even a very slight change in temperature in the volume of a liquid when it is near its critical temperature was observed in 1852 by Waterston during experiments in glass tubes.[1] The rate of diminution in volume must, however, fall off rapidly, since the diminution in amplitude of molecular movement for a given diminution in translational energy will fall off quickly. With further fall in temperature the diminution in amplitude of molecular movement must soon reach a point at which the diminution in volume is only proportional to the diminution in translational energy, as will easily be seen from Fig. 7. This signifies that in the equation $(p+\pi N^2)(1-Nb)=NRT$ the rate of diminution in the value of π has become proportional to the rate of increase in the value of N.

At the same time the rate of increase in the density of the liquid per degree of fall in temperature will have become much slower. The value of p (the vapour pressure) will also have become a mere fraction of that of πN^2 (the intrinsic pressure), since, though they were equal at the critical temperature, πN^2 will have increased very greatly at the expense of p with the fall in temperature and very rapid increase of N.

---

[1] Waterston's *Collected Scientific Papers*, p. 513.

Thus the total internal pressure will soon become nearly the same as the intrinsic pressure. The equation for a liquid not under artificial pressure will thus become practically $\pi N^2(1 - Nb) = NRT$, or $\pi N(1 - Nb) = RT$; and so long as π continues to vary inversely as N, so that πN is a constant, the intrinsic pressure will only increase with N as the temperature falls. Both theoretical and experimental evidence that πN remains a constant over a wide range of temperature will be adduced below.

Before carrying the argument further it is necessary to discuss the question of molecular compressibility. That molecules or atoms are relatively incompressible is quite evident from the relations of pressure to gross volume in permanent gases at increasing pressures; but the theory which has so far been adopted above assumes that the incompressibility is no more than relative, and that compressibility must become quite sensible at a high pressure. Nevertheless there is considerable apparent evidence that under the external pressures actually measured in gases and liquids the molecules are almost, if not completely, incompressible. Before discussing this evidence we may consider what can be anticipated as regards molecular compressibility from the theory embodied in Fig. 7.

When there is pressure in a collection of molecules such as a portion of liquid or gas, this is equivalent in a statistical sense to an increase in the attractive force which acts between the molecules. We can thus represent the virtually increased attractive force by a line such as the dotted line in Fig. 8. It will be seen that with the increased pressure the point of intersection of the two lines has been shifted inwards towards the centre of the molecule. The virtual volume of the molecule is thus correspondingly decreased. The angle of intersection of the two lines has also become blunter, which of course signifies that the resistance to further compression has been increased. It will also be seen that the increase in kinetic energy as two molecules approach one another from a given distance between their centres is increased; and this is a matter of vital importance in its bearing on the internal pressure in a liquid.

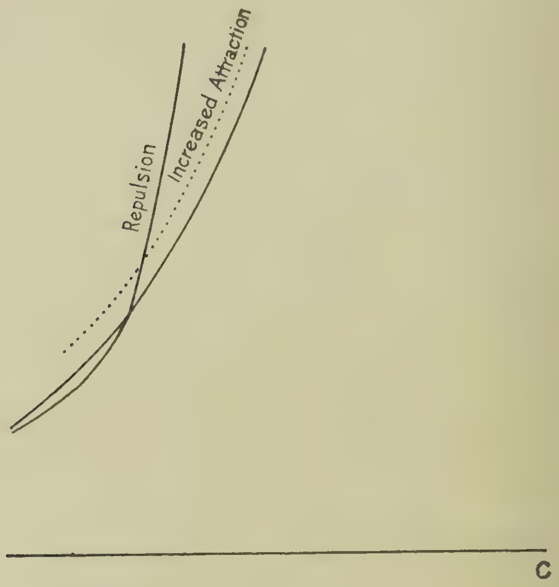

FIG. 8.

The attractive force represented by the continuous line is, according to the theory, very great as compared with net molecular or gravitational attractive force. It can only, therefore, be with very heavy increases of pressure that there is any sensible diminution in the volume of the molecules; but in a liquid at well below its critical temperature the internal pressure is enormous, and various facts point to the conclusion that this internal pressure is sufficient to produce a quite sensible diminution in molecular volume, which depends on the pressure, so that the virtual volume occupied by the molecules diminishes sensibly in direct proportion to increase in the total pressure. As a corollary to this conclusion it also follows that when, in cooling down below its critical temperature, a liquid reaches the point where π diminishes in proportion to N, this rate of diminution in the value of π may continue with further drop in temperature, since, in proportion to the further molecular compression, the kinetic energy acquired by a molecule in approaching another molecule from a given distance may increase, so that π will diminish in proportion as N increases, with the result that πN continues to remain a constant. This conclusion or assumption, as we shall presently see, furnishes a key to a large group of facts relating to liquids; but meanwhile we may consider certain facts which are generally regarded as being inconsistent with the theory of sensible molecular compressibility.

We may first take the facts embodied in the well-known Ramsay-Young law. It was discovered by Ramsay and Young,[1] and confirmed by Young for several liquids and gases, that when the liquid or gas is kept at constant volume and the temperature is raised, the external pressure rises in very exact proportion to the rise in temperature, so that the pressures when plotted against temperatures give straight lines. This would be the case if the value of b remained constant in the fundamental equation $P(V = b) = RT$, or the corresponding equation with the new notation, $P(1 - Nb) = NRT$. If, on the other hand, the value of b varied in any inverse proportion to the external pressure, the plotted lines would be curved. Hence the Ramsay-Young results are generally considered to show that there is no sensible compression of the molecules as the pressure rises.

In reality the Ramsay-Young law is consistent, not only with the theory that the molecules are not appreciably compressed, but also with the theory that they are compressed in the ratio, for either a liquid or a gas, of P the whole initial internal pressure to $P + p$ the final internal pressure, just as the theory enunciated above assumes. On either theory the pressure, when plotted against temperature, gives a straight line when N remains constant; and since the theory that molecules are rigid incompressible bodies is inconsistent with modern conceptions of what a molecule or atom is, we must, I think, prefer the alternative that molecules are compressible in proportion to the total internal pressure.

If, as in the case of water, N varies owing to dissociation of solid aggregates by heat (which, to judge from Waterston's experiments with water, would seem to be the case up to nearly 200° C.) the Ramsay-Young law does not hold. It seems to be only when there is no appreciable molecular dissociation or association of any kind that this striking law does hold.

---

[1] Ramsay and Young, *Phil. Mag.*, 1887.

The success with which the van der Waals' equation, in which *b* is treated as a constant, enables us to predict the changes in volume or pressure of a gas at constant temperature, but widely varying pressure or volume, has also been taken as evidence that molecules are not sensibly compressible under ordinary attainable pressures. If the equation held good at very high pressures, this would certainly indicate that sensible compression of molecules does not occur. But the equation ceases to correspond with observed results at very high pressures; and in any case we have already (p. 227) seen reason to conclude that the actual diminutions in both the *a* and *b* of van der Waals' equation tend to mask one another in calculations from the equation. Apart from the physical theory which tells us why van der Waals' equation holds so closely within certain limits, but not further, the equation itself is quite unintelligible. Van der Waals himself concluded that in a gas the actual volume of the molecules, as indicated by the volumes of liquids and solids at far below the critical temperature, is about one-fourth the volume of *b*. His suggested explanation of the difference is, however, very unconvincing, and depends on the unintelligible assumptions that molecules are rigid bodies, and that the *a* of his equation remains constant.

The known deviations from what may be called the typical law of compressibility of liquids may be accounted for by the interference brought about through the presence of fleeting solid aggregates in liquids. Thus owing to the presence of solid aggregates the compressibility of water is very abnormal, especially at temperatures near the freezing-point.[1] Water also does not conform to the Ramsay-Young law. This is due to the fleeting formation of bulky solid aggregates being diminished by the rise of pressure or temperature. Ice itself occupies a gross volume about 8 per cent. greater than that of an equal mass of water. Other liquids may be abnormally compressible owing to progressive formation of solid aggregates which are less bulky than the corresponding liquid.

From the above discussion it will be seen that so far the facts are consistent with the theory that from a point not much below the critical temperature the value of πN in a liquid remains constant at lower temperatures. We must now consider further evidence in this direction.

If, in addition to the natural internal intrinsic pressure of a liquid, pressure is applied from without, the further diminution in the virtual volume of the molecules will, on the theory under discussion, be proportional to the ratio of the new pressure to the natural internal pressure. If we call the latter pressure P, and the added pressure P_0, the diminution will be proportional to $\frac{P+P_0}{P}$. Now this is just the diminution required to make the osmotic pressure equation $P_0 = \frac{nN}{N_1 - n}\,{\cdot}082T$ accord exactly with the observed values up to about 70 atmospheres, according to the very exact experimental results of Berkeley and Hartley, as given in Chapter IV. Here, then, we have further confirmation of the theory put forward as to the rate of variation of b with pressure, and at the same time the theoretical justification for the exact equation given for osmotic pressure.

---

[1] See, for instance, the graphs published by Amagat, showing the results of compression to 3000 atmospheres of water and ether, in *La Statique des fluides*, 1917.

Owing to the enormous internal pressures which, according to the theory which has been developed, must exist in the liquid form of substances for which the value of π in their gas equation $(p+\pi N^2)(1-Nb)=NRT$ (or the value of a in van der Waals' equation) is high, the value of $1-Nb$ in the liquid form must be a very small fraction of unity, so that as the temperature rises (provided that no molecular dissociation occurs) the value of N can only rise very slowly. Such substances, therefore, since their intrinsic pressure, according to the theory, only diminishes as N does, and N diminishes very slowly until their critical temperatures are not very far off, ought to behave, as regards expansion with temperature, like gas under constant pressure, except that the pressure is diminishing very slowly with rise of temperature. Now this is just how a number of liquids, all of which can be used in thermometers, actually do behave. The rate at which they expand does increase with rise of temperature, so that they are never quite comparable to a gas which follows Charles's law; but the rate of increase is very slow in these liquids, and this is particularly the case with mercury. Thus when two fixed points, such as the freezing and boiling temperatures of water, have been marked on a mercury thermometer, and the bore is even, intermediate temperatures are indicated with almost exact accuracy by the thermometer if the stem between the two fixed points, and beyond them, is evenly divided. The fact that liquids (as well as solids) exist which can be applied in a simple manner for measuring temperature affords further distinct confirmation of the theory that πN becomes a constant when the temperature of a liquid has fallen to a short way below the critical temperature. If it were otherwise, liquids could not be used in thermometers without great difficulty in correct graduation.

It is of course only a normal liquid (*i.e.* a liquid which is not, like water, full of dissociable solid aggregates), and only a liquid of very high internal pressure, that can be used in an ordinary thermometer. If the liquid has a low internal pressure (in other words a low critical temperature), its rate of expansion increases too rapidly with temperature, so that there are serious errors in the indications which it gives between the fixed points, if the stem is evenly divided.

As the temperature of a liquid gets nearer and nearer to its critical point its expansion per degree of rise of temperature must, according to the theory which has been developed, become greater and greater; and this is in accordance with observation. Waterston, for instance, relates that in observing the expansion of liquids at near their critical temperatures, he was much troubled by the level of the liquid oscillating up and down with the smallest change of temperature. Between 320° and 374° (its critical temperature) water doubles its volume, as Ramsay and Young found, while a gas under constant pressure would only expand by about 9 per cent.

When the temperature has risen to pretty close to the critical temperature the value of πN must, according to the theory, diminish rapidly, so that the increase in volume at still higher temperatures becomes specially rapid until the substance ceases to be a liquid when the critical temperature is reached.

Great attention was given by Waterston to the normal law of liquid expansion. He concluded, in his first paper on liquid expansion (*Collected Papers*, p. 365), that this law may be expressed by the equation $\left(\frac{\gamma - t}{k}\right)^{\rho} = \Delta$, where γ represents a temperature which is a little above the critical temperature (called by Waterston the "point of transition") of the particular liquid, Δ is the density of the liquid, and ρ is a constant which must be determined from two or more observations of the density of the liquid at different temperatures. The liquid continues to expand (or reduce its density) in accordance with this equation till it comes close to the critical temperature, when it must then expand much more rapidly since it cannot reach the temperature γ in the liquid state.

The relation of the expansion of a normal liquid to its critical temperature is thus evident, together with the fact that the closer the approach to the critical temperature the greater the rate of expansion with rise of temperature becomes. The regularity of the increase in rate of expansion with temperature corresponds to the regularity with which the value πN^2 diminishes with temperature and the value of N until a point close to the critical temperature is reached. This is a further indication that over a very wide range of temperature, from near the critical temperature downwards, the value of πN is constant, so that the intrinsic pressure, πN^2, diminishes with rise of temperature simply in proportion to N, or in proportion to the diminution of density.

Waterston devoted a special paper to his investigations on the expansion of water at different temperatures. As was already very well known, the expansion of water with temperature is at ordinary temperatures completely abnormal, and below 4° water expands on cooling, owing to rapid increase in momentary formation of solid aggregates. He succeeded in following the expansion of water up to 330° C., which is not far from the critical temperature (now known to be about 374°), and found, to his great satisfaction, that above 190°, though not below, the expansion of water had become normal. The readings above 190° gave a value for γ of 429° C., which is, as could be expected from the behaviour of other liquids, above the critical temperature. Thus at over 190° the fugitive solid aggregates are no longer present in water; but that they are still present even at temperatures up to nearly 200° is clearly indicated also by the want of correspondence of water with the Eötvös law, which will be referred to presently.

The expansion of a normal liquid in presence of its vapour has been treated as simply determined by progressive diminution of intrinsic pressure, πN^2, as distinguished from total internal pressure. This is because it is only the intrinsic pressure that counts. The rest of the internal pressure is the same inside and outside the liquid.

We may now consider the distribution of internal pressure at the surface of a liquid, and the phenomena of what is now usually called "surface tension." The intrinsic pressure (*i.e.*, the pressure dependent on molecular attraction) must, according to the dynamical theory of gases, be πN^2, and the value of N in the interior of a considerable mass of liquid is easily obtained from the specific gravity and molecular weight of the liquid substance. It is different, however, in the immediate neighbourhood of the free surface of the liquid, and that there must be a difference in pressure was clearly evident to Laplace, though of course he did not regard the phenomena from the standpoint of energy.

The layer of molecules next the surface will be violently bombarded from below, but hardly at all from above. They will thus be free to travel much further in a direction at right angles to the surface, so that their mean position will be shifted outwards and they will be further apart from one another than are the molecules which are deep down in the body of the liquid. But the consequence will be that the layer of molecules next behind them, since the bombardment from above will be lessened, will also be lifted outwards, though to a less extent, since the pressure from outside will now be less than from within. And so the process will proceed to deeper and deeper layers, until at last the pressure is sensibly the same as in the interior of the liquid. Since the molecules are further apart, on the whole, than beneath the surface layers, the value of N will be less in the surface layers, and consequently the pressure will be less.

Just as "osmotic pressure" is commonly described as a positive pressure due to a cause inherent in a liquid, so "surface tension" is sometimes described as an extra surface pressure tending to compress a liquid. But actually, as we have just seen, the pressure must be less in the surface layers than beneath them. Why, then, do small collections of liquid tend to assume the form of drops? and why do liquid films resist extension?

We can see the reason at once if we consider what must be happening when the surface layer of liquid is being extended or reduced. In the extension of a liquid surface the surface molecules are being separated against the force of molecular attraction, while in the reduction of a liquid surface the surface molecules are being brought together under the influence of molecular attraction. When a surface is being reduced they are thus taking up a natural position, just as a stone suspended in the air takes up a natural position when it falls to the ground. Heat is, moreover, produced at the expense of potential energy when a surface is reduced, just as heat is produced when a stone falls to the ground. Conversely, work is done against the force of molecular attraction in the surface molecules if the surface of a liquid is extended, and in this process heat disappears, or becomes "latent" in so far as work is done at the expense of the heat energy of the liquid. The work done when a liquid surface extends is in separating molecules against the force of molecular attraction, just as when the liquid is evaporated. Hence this work, as Waterston seems to have been the first to point out,[1] must vary in some essential connection with the latent heat of evaporation.

We can measure the force by measuring the force required per unit breadth of film (with its two surfaces), to extend a film of liquid; but for comparative purposes the measurement is more conveniently made indirectly by measuring the height to which a column of liquid is raised in a glass tube of very narrow and quite even bore, when the liquid wets the glass. From this we can calculate the "surface tension," and all the measurements, by Ramsay and Shields and others, indicate that it diminishes parallel with increase in temperature, up to a temperature not far from the

[1] Waterston, *Collected Scientific Papers*, p. 407.

critical temperature, when it begins to diminish more and more slowly till zero is reached at the critical temperature. The latent heat of evaporation, like the value of N, diminishes at an increasing rate with rise of temperature, till it reaches zero at the critical temperature, and appears to diminish with N till near the critical temperature. The essential connection between surface tension (or "capillarity" as Waterston called it), latent heat of evaporation, and expansion, as well as increase in vapour density, was clearly pointed out in Waterston's disregarded papers.

A diminution in the latent heat of evaporation with increase in the value of N for a normal liquid up to near the critical temperature can be interpreted easily on the theory that within wide limits of temperature the value of πN is constant in the equation of state $(p+\pi N^2)(1-Nb)=NRT$. So long as πN is constant the work done in separating a given number of molecules in a liquid must diminish as N diminishes in accordance with the theory embodied in Figs. 7 and 8. The internal work expended in separating the molecules per unit surface of a liquid when the surface is extended will therefore diminish in direct proportion to rise in temperature. Hence the empirically discovered connections between rise of temperature, diminution of latent heat of evaporation, diminution of surface tension, and expansion or increase in molecular volume, become intelligible, and a further reason is added in favour of the conclusion that πN remains constant over a very large range of temperature.

We can now discuss the relation between surface tension or energy and molecular volume or size of the liquid as its temperature rises. Assuming that the molecules in a liquid are free, and not partly aggregated together, the value of N represents the molar concentration of free molecules per litre. As the value of N diminishes with rise of temperature the number of molecules per unit of surface area will diminish as $N^{\frac{2}{3}}$. It was partly shown by Eötvös,[1] and completely by Ramsay and Shields[2] for a number of liquids, that if we call σ the "surface tension" of the liquid, T_c the critical temperature, and T the temperature of the liquid, then, adopting the meaning given to N in this book, σ varies as $2.1\,(T_c - 6 - T)\,N^{\frac{2}{3}}$, up to near the critical temperature.

In the case of certain liquids, such as water, the law does not hold if we calculate the value of N from the total concentration of molecules, regardless of whether they are free or more or less aggregated together. But, as already pointed out, we have abundant independent evidence of molecular aggregation in the case of water; and the same is true of other liquids to which the Eötvös law does not apply. The deviation or non-deviation of a liquid from this law gives us, in fact, clear information as to how far the molecules of the liquid are aggregated together within any range of temperature, as Ramsay and Shields maintained. Water at a temperature above boiling-point is known to give results deviating less from the equation than at lower temperatures; but it would seem from Waterston's results, already alluded to, for the expansion of water that constancy cannot be reached below 190° C.

---

[1] *Annalen der Physik.*, 1886.

[2] *Trans. Chem. Soc.*, 1893.

What is the significance of the constant 2.1 in the Eötvos-Ramsay-Shields equation? The equation indicates clearly that surface tension is the force required to separate a single layer of molecules beneath the surface of the liquid. When the attraction holding them is overcome, the surface is extended, with the necessary accompaniment that internal work is done by the liquid in proportion to the increase in extension of the surface.

But if this is the meaning of surface tension it follows at once that a gas at the critical temperature has still a considerable surface tension, while the equation as it stands gives only the liquid surface tension. At the critical temperature and pressure, as we have seen (page 237), the attractive forces, represented by the value of πN^2, are just sufficient to balance the disruptive forces represented by the value of T_c. The equation signifies that if πN^2 continued to diminish as N with rise of temperature the surface tension of the gas would be equal to that of the liquid at $T_c - 6$. The measurable liquid surface tension would thus be extinguished. The substance would, however, still have a true surface tension varying as $N^{\frac{2}{3}}$ and as the ratio of $T_c - 6$ to T, and its value would, from a value of unity at $T_c - 6$, increase in proportion as T was below $T_c - 6$. In other words, the true surface tension of the substance would vary as $N^{\frac{2}{3}} \frac{T_c - 6}{T}$, while the measurable surface tension would vary as $N^{\frac{2}{3}} (T_c - 6 - T)$.

Since at $T_c - 6$ the surface tension of the gas would be equal to that of the liquid, the surface tension of the substance would be twice the true surface tension of either the gas or the liquid. The unit value for the substance would thus be twice the true value of unit liquid surface tension. If, therefore, we take the true surface tension of the substance as unity, we have to multiply the true surface tension of the liquid by 2 in order to express its value in terms of unity. Its measured value would thus be $2N^{\frac{2}{3}} (T_c - 6 - T)$.

The actual value by the Eötvös-Ramsay-Shields equation is $2.1 \ N^{\frac{2}{3}} (T_c - 6 - T)$. The reason why the factor is 2.1, and not 2, seems to be that the actual critical temperature is T_c, and not $T_c - 6$. The consequence of this is that the true surface tension of the liquid is somewhat less than it would be if πN remained constant up to the critical temperature. The reason why it is not constant till well below the critical temperature has already been discussed (p. 246).

It may at first sight seem difficult to attribute a surface tension to gases. In reality, however, this is no more difficult than to attribute a surface tension to solids, though in either case we have no direct means of measuring the surface tension. If, however, we know the critical temperature and the actual temperature of a pure liquid substance, and the true value for it of N, we can predict its surface tension.

Since over a large range of temperature up to not far from the critical temperature the surface tension of a liquid diminishes in proportion to rise of temperature, it follows that over this range the rate of diminution of $N^{\frac{2}{3}}$ is such as to make this possible. The value of $N^{\frac{2}{3}}$ is, however, governed by that of πN^2, the intrinsic pressure, and over the same range of temperature πN^2 diminishes as N diminishes. The diminution of surface tension with temperature is consequently even over the range, and far greater than in proportion to rise in absolute temperature.

Surface tension is sometimes referred to as if it were a mysterious force capable of accounting for all sorts of facts not otherwise accounted for. In reality it is thoroughly intelligible from a physical standpoint, and the limitations in what it will account for are quite definite.

The manner in which the presence of association among the molecules of a pure liquid affects its surface tension furnishes a key to the influence of solutes on surface tension. Solutes are apt, by the process of adsorption already alluded to, to become concentrated in the surface layers of the solvent and in this way sometimes reduce its surface tension very greatly. A well-known example is the influence of bile-salts in almost abolishing the surface tension of water. The solute may also solidify where it is concentrated, as in the case of protein solutions and other solutions recently studied by Ramsden. The phenomena observed at the surface of contaminated mercury afford another interesting instance.

Consideration must now be given to the internal pressures of solutions. The phenomena of osmotic pressure, and of depression of freezing-point, depression of vapour pressure, and rise of boiling-point in solutions, are alike in indicating, as we have seen, that up to certain concentrations, solutions have the same intrinsic pressure as the pure solvents at the same temperature. Why should this be so? In different liquids, for example, the intrinsic pressures, as indicated, for instance, by their different critical temperatures, vary greatly. Yet the depressions of vapour pressure of the solvent, when one liquid is dissolved in another, do not, up to certain concentrations, indicate any difference in intrinsic pressure between pure liquid and solution.

Just as the liquid state itself is dependent on the existence of mutual attraction between the molecules of the liquid substance, so, evidently, is the existence of solutions (in which the solute is present in the liquid state) dependent on mutual attractions between solute and solvent molecules. Gases are miscible in all proportions, since their molecules are almost indifferent to one another; but solutes may only be present in certain maximum proportions at any given temperature; and the limited solubility of solutes must be dependent on the limited attraction between solute and solvent molecules as compared with the attraction of solvent or solute molecules for one another. The molecules of ordinary gases, for example, as well as of many liquids and solids, have no appreciable attraction for the molecules of mercury, and are therefore quite insoluble in this liquid, though various metals dissolve readily in mercury. Any gas molecules which penetrate between the surface molecules of the mercury will be literally squeezed out owing to the enormously greater attraction of the mercury molecules for one another than for the gas molecules.

The adsorption of gases or liquids on the surface of solids, or on internal surfaces when the solids are porous, is likewise due to attraction between the gaseous or liquid molecules and free molecular surfaces of the solid. Adsorption is thus a process which is essentially similar to solution, though only the free molecular surfaces of the adsorbing solid or liquid are concerned, and therefore only a small part of the molecules present in the solid or liquid. Liquid adsorption, or concentration of foreign molecules on the surface of a liquid, is as widespread a phenomenon as solid adsorption. The great practical importance of both solid and liquid adsorption is becoming constantly more evident. The surface concentration of solutes and the formation of liquid surface films are examples of adsorption.

In considering the phenomena of solutions we have primarily three factors to consider: the attraction of solvent for solute molecules; of solvent molecules for one another; and of solute molecules for one another. We may also have to take into account the formation of compound molecules such as hydrates, of solute and solvent, the splitting up of solute molecules in the formation of ions, and the formation of compounds of solute or solvent molecules with one another in the solution. Hydration or "solvation," and ionisation, have already been referred to frequently. These secondary phenomena need not, however, be considered in a general discussion of the primary factors in solution.

Solutions of gases in liquids may be considered first, as being most easily intelligible on the dynamical theory of matter. The reason why a permanent gas retains the gaseous form at ordinary temperatures and pressures is that the molecules of the gas have so little attraction for one another. In other words, the value of π in the equation of state is so low that the molecules do not condense to the liquid or solid form. In this respect they are a contrast to molecules of other substances for which the value of π is so high that condensation occurs readily, and a still greater contrast to molecules with a value of π so high that the substance remains solid at ordinary temperatures, and may even do so, like carbon and certain other substances, at a white heat. As a general rule molecules of the permanent gases have also but slight attractions for the molecules of liquids, and correspondingly low solubilities.

It has been pointed out at p. 191 that when a liquid is saturated with a gas or vapour a state of equilibrium exists in which, if p is the pressure of the gas or vapour, and $1 - N_1b_1$ represents the proportion of free space in the liquid, $(1 - N_1b_1)\,p$ represents the diffusion pressure of the gas or vapour into or out of the solution. In accordance with the dynamical theory of matter the value of $1 - N_1 b_1$ varies inversely as that of N_1, the total molar concentration of the solution, so long as the internal pressure P of the liquid remains the same in spite of the addition of solute molecules, and the vapour pressure of the solute varies as $\frac{n}{N_1}$. By similar reasoning the vapour pressure of the solvent varies as $\frac{N_1 - n}{N_1}$, or $1 - \frac{n}{N_1}$. This would no longer be the case, however, if P varied with increasing concentration of the solute.

Now the facts discovered by Raoult with regard to the vapour pressures of solvents, as given in the table on p. 186, correspond to the conclusion that the vapour pressure of the solvent was actually varying as $1 - \frac{n}{N_1}$, and therefore that the internal pressure of the solution was remaining steady in spite of the presence of the solute. The latter conclusion follows also from the data given as to osmotic pressures and depressions of freezing-points in sugar solutions up to high concentrations, and as to the rises in boiling-points of sugar solutions up to a concentration of nearly 20 grms. per 100 grms. of water.

Let us now, however, examine the rest of Raoult's data as to the depressions in the vapour pressure of ether in solutions in that solvent. Only those for moderate concentrations were given in the table on p. 186. The solutes used by Raoult were miscible in all proportions with the solvent, so that he was able to obtain results up to the highest concentrations. The results are shown in the appended table.

Results of Observations of Depression of Vapour Pressure of Ether in Solutions in it.

OIL OF TURPENTINE (B.P. = 160°), IN ETHER (B.P. = 35°).

Percentage of Solute by Weight.	$\frac{n}{N_1}100.$	$\frac{d}{p}100$	$\frac{n}{N_1}k\,100$	
10·0	5·9	6·0	5·3	$k = 0·90$
20·2	12·1	11·9	10·9	
35·9	23·4	21·9	21·1	
50·8	35·5	32·4	32·0	
62·8	47·9	43·8	43·1	
76·9	64·5	57·9	58·0	

NITROBENZOL (B.P. = 205°) IN ETHER.

Percentage of Solute by Weight.	$\frac{n}{N_1}100.$	$\frac{d}{p}100$	$\frac{n}{N_1}k\,100$	
9·6	6·0	5·5	4·4	$k = 0·74$
26·7	17·9	14·2	13·2	
47·2	35·5	25·6	26·3	
65·4	53·2	38·0	39·4	
80·9	75·9	55·6	56·2	
88·5	84·0	64·5	62·2	

ANILIN (B.P. = 182°) IN ETHER.

Percentage of Solute by Weight.	$\frac{n}{N_1}100.$	$\frac{d}{p}100$	$\frac{n}{N_1}k\,100$	
4·8	3·85	4·0	3·4	$k = 0·90$
9·5	7·7	8·1	6·9	
18·1	14·8	15·4	13·3	
24·5	20·5	19·7	18·4	
55·3	44·6	42·4	40·1	
73·4	68·7	57·6	61·8	

METHYL SALICYLATE (B.P. = 222°) IN ETHER.

Percentage of Solute by Weight.	$\frac{n}{N_1}100.$	$\frac{d}{p}100$	$\frac{n}{N_1}k\,100$	
2·26	1·1	0·4	0·9	$k = 0·82$
4·2	2·1	1·7	1·7	
9·4	4·1	4·0	3·4	
17·3	9·2	8·6	7·5	
26·8	15·1	13·0	12·5	
38·6	23·2	18·9	19·0	
66·4	49·0	40·6	40·2	
87·3	77·0	63·9	63·1	
91·4	85·0	70·8	69·7	

ETHYL BENZOATE (B.P. = 213°) IN ETHER.

Percentage of Solute by Weight.	$\frac{n}{N_1}100.$	$\frac{d}{p}100$	$\frac{n}{N_1}k\,100$	
9·4	4·9	5·1	4·4	$k = 0·90$
17·7	9·6	9·1	8·6	
43·0	27·1	24·8	24·4	
69·6	52·3	47·1	47·1	
86·2	75·5	70·0	67·9	
97·1	94·4	87·6	85·0	

It will be seen from this table that with the higher concentrations the relative depression of vapour pression is nothing like proportional to $\frac{n}{N_1 - n}$, the ratio of solute to solvent molecules, which determines osmotic pression, depression of melting-point, and rise of boiling-point. Except for concentrations below about 20 per cent. it also diverges very distinctly from the ratio $\frac{n}{N_1}$, that of moles of solute to total moles. It is, however, given approximately by multiplying the ratio $\frac{n}{N_1}$, by a constant, which is below unity and varies from 0.90 to 0.74 in the five cases investigated. Raoult concluded that it is always necessary to multiply the ratio $\frac{n}{N_1}$ (which is equivalent to the N of his own notation as applied to depressions of vapour pressure) by a suitable constant in each case. When, however, we examine the figures closely, we see that up to concentrations of about 20 per cent. the results agree better with the calculation if no constant is used.

In every case the solutes have a much higher boiling-point than the solvent. We should therefore expect that for strong solutions the internal pressure of the solution would rise, and therefore give a smaller depression of vapour pressure than proportional to $\frac{n}{N_1}$. This is quite evidently the case, but up to a considerable concentration the depression is proportional to $\frac{n}{N_1}$, in accordance with the theory already discussed. The diminished rate of depression at high concentrations has evidently the same cause as the increased rate of rise of boiling-point in sugar solutions, as shown in the table at p. 204. We obtain a much clearer insight into this matter by examining data obtained at a later date by v. Zawidzki,[1] and plotted graphically. He obtained the vapour pressures of both solute and solvent in a large number of solutions of organic substances in one another. The depressions of vapour pressure are plotted against the values of $\frac{n}{N_1}$, or the "molar fraction."

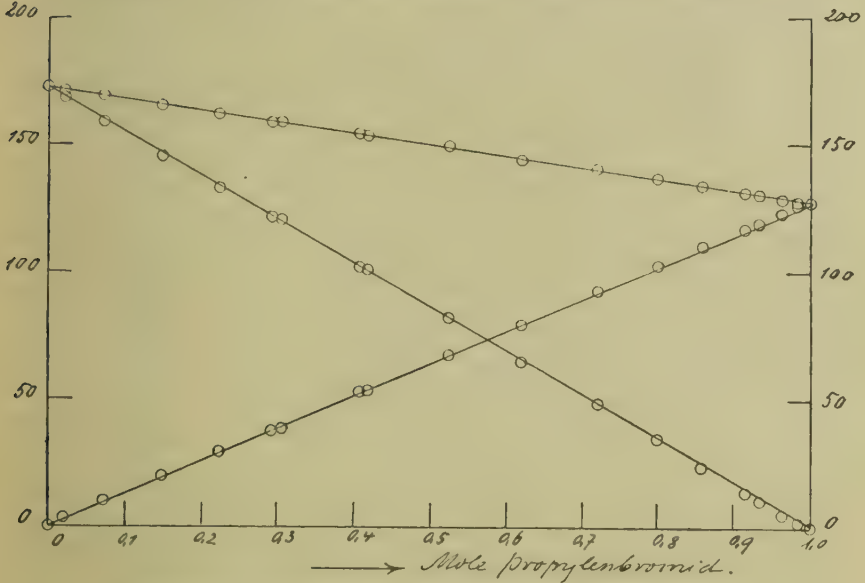

FIG. 9.—Propylene Bromide in Ethylene Bromide. *After* v. Zawidzki.

---

[1] Von Zawidzki, *Zeitschr. f. Physikalische Chemie*, 1900, xxxv., p. 129.

In two cases (solutions of propylene bromide in ethylene bromide and of benzol in ethylene chloride) the graphs were straight lines (see Fig. 9), so that the depression of vapour pressure of both solute and solvent varied directly as the value for either solute or solvent of $\frac{n}{N_1}$ (Raoult's N). This means that it was quite indifferent whether either substance in each pair was called "solute" or "solvent." The boiling-points of propylene bromide and ethylene bromide are 141.6° and 130°; those of benzol and

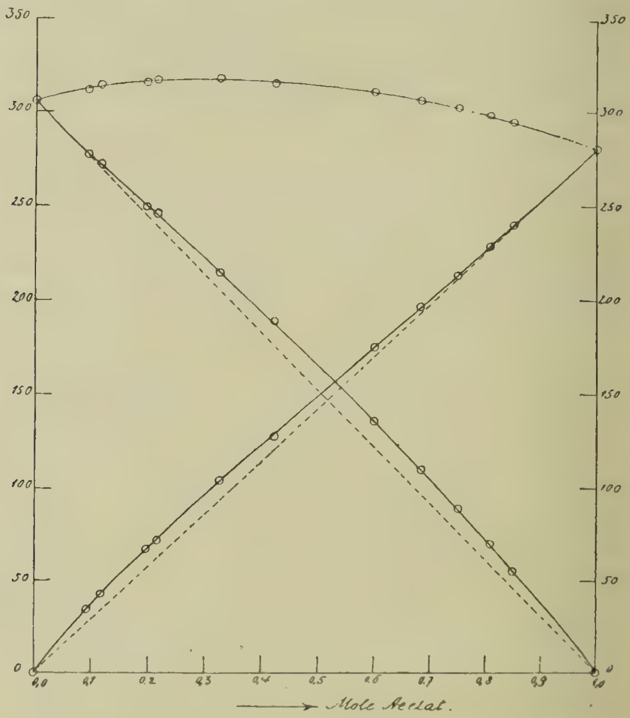

FIG. 10.—Carbon Tetrachloride and Ethyl Acetate. *After* v. Zawidzki.

ethylene chloride are 83.4° and 83.4°. Thus in each pair the boiling-points are very close, and so also

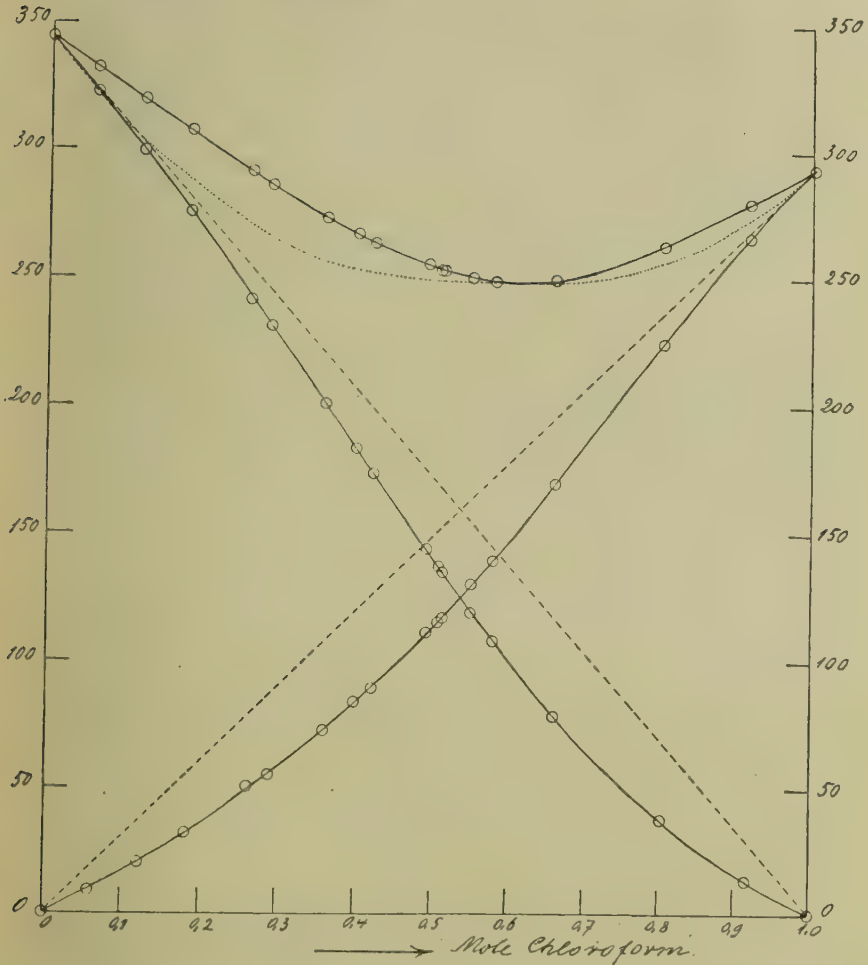

FIG. 11.—Acetone and Chloroform. *After* v. Zawidzki.

would the critical temperatures doubtless be if they could be determined.

In the great majority of the cases the complete graphs were not straight lines or even near to being straight lines, but wherever a sufficient number of points were given to show the precise forms of the graphs, it appears that for some distance at either end the graphs were straight lines. This is shown in Figs. 10 and 11. So long as the graph for any solute continues to be a straight line we can consider that the solute is dissolved in the solvent, while the solvent is dissolved in the solute when the graph again becomes straight. While the graph is curving, we cannot regard the solution as being either in the solvent or the solute. In cases where the graph becomes curved there is a considerable difference in the boiling-points of solute and solvent, though the matter is sometimes complicated by the presence of molecular combinations in solute or solvent.

Accurate data as to depression of vapour pressure and increase of osmotic pressure with cane-sugar solutions of very high molar concentration have been obtained by Berkeley, Hartley, and Burton.[1] They not only determined the depressions of vapour pressure at 0° and 30° up to concentrations of 960 grammes of anhydrous cane-sugar per litre, but proved experimentally that Porter's equation,[2] which connects relative depression of vapour pressure with osmotic pressure, gives extremely accurately the osmotic pressure when the relative depression of vapour pressure is known, or *vice versa*, though it does not connect either vapour pressure or osmotic pressure with molar concentration. By means of this equation they were thus enabled to obtain the osmotic pressures for extremely concentrated solutions from actual observations of the relative depressions of vapour pressure, though the experimental difficulties in obtaining direct observations of osmotic pressure were too great.

---

[1] Berkeley, Hartley, and Burton, *Phil. Trans.*, A, 218, p. 295, 1919.

[2] Porter, *Proc. Royal Society*, A, 80, p. 460, 1908.

The accompanying table, in which columns 1, 4, and 7 are taken directly from their paper, has been calculated from their results and from corresponding specific gravity data.

Grammes Anhydrous Sucrose per 100 grammes Total Water.	Specific Gravity.*	Grammes Anhydrous Sucrose per 100 c.c.	Observed or Calculated Osmotic Pressure in Atmospheres.	Observed or Calculated Osmotic Pressure divided by ·082 NT.	Moles Sucrose Pentahydrate to Moles of Free Water, $\frac{n}{N_1 - n}$.	Observed Relative Depression of Vapour Pressure of Water, $\frac{pd}{p}$.	Moles Sucrose Pentahydrate to Moles of Solution, $\frac{n}{N_1}$.	Moles Sucrose per Litre of Solution to Moles Water per Litre of Pure Water, $\frac{n}{N}$.
34·52	1·118	28·7	24·91	·0201	·0202	·0196‡	·0198	·0151
56·5	1·165	42·2	43·84	·0353	·0356	·0346	·0344	·0222
81·2	1·207	54·1	67·68	·0545	·0560	·0524	·0530	·0285
112·0	1·258	66·4	100·43	·0808	·0875	·0766	·0805	·0350
141·0	1·305	76·6	134·71	·1085	·126	·1013	·112	·0404
183·0	1·330	86·0	186·86†	·1505	·198	·1365	·165	·0453
217·5	1·345	92·1	230·7†	·1857	·314	·1656	·235	·0485
243·0	1·357	96·1	264·46†	·2129	·435	·1855	·303	·0506

* Calculated from table quoted by Landolt and Börnstein.
† Calculated by Porter's equation.
‡ From data of Smits (page 187, above).

It will be seen from this table, and those previously given at pages 146, 149, 186 and 187, that although up to a concentration of about 50 grammes per 100 c.c. the osmotic pressure varies very exactly as $\frac{n}{N_1 - n}$, and the vapour pressure as $\frac{n}{N_1}$, assuming that the sugar is present as pentahydrate, at higher concentrations both the osmotic pressure and vapour pressure give values increasingly lower than in these proportions. This might be partly due to diminishing hydration of the sugar, but is probably due entirely, or almost entirely, to increasing intrinsic pressure in the solutions, just as in solutions of one liquid in another, where the solute has a higher critical temperature or boiling-point than the solvent.

It is very clear from the table that depression of vapour pressure does not increase parallel with increase in osmotic pressure, and still more clear that there is no parallelism between depression of vapour pressure and mere increase in the molar concentration of the solvent.

There seems, thus, to be no difficulty in following the changes of internal pressure which are finally produced in solvents in consequence of the presence of solutes. As soon as we think of what must be happening when an indefinitely large proportion of solute molecules are mingled with a solvent, it becomes evident that unless the attraction of solute molecules for one another and for the solvent molecules are equal, the internal pressure of the solution cannot possibly remain constant with an indefinitely large proportion of solute; and Raoult's and Zawidzki's results indicate quite clearly that when the proportion of solute exceeds in each particular case a certain amount the internal pressure begins to change. The same fact is indicated by the manner in which the rises of boiling-point of very strong sugar solutions and of osmotic pressures and depressions of freezing-point behave (see tables at pages 204, 275, 178). But why does the internal pressure remain constant up to a considerable concentration of solute?

In a solid solute, or liquid solute of higher critical temperature or boiling-point, the solute molecules must have greater attraction for one another than the solvent molecules, so that the internal pressure of the pure solute must be higher than that of the solvent. Let us consider what must happen when the solute is dissolved in the solvent, the molecules of which have presumably less attraction for one another than the solute molecules.

If solute molecules have more attraction for solvent molecules than the solvent molecules have for one another, so that the value of π as between solvent molecules is less than as between solute and solvent molecules, it would seem at first as if the mean internal pressure in the solution must be greater than in the pure solvent. To put the matter into different language, the mean volume of clear space round each solute molecule would apparently be less than round each molecule of the pure solvent; hence the mean clear space per molecule would be less in the solution than in the pure solvent.

The reason why this is not so till the concentration becomes very considerable appears to be as follows. When solute molecules are more attracted by solvent molecules than solvent molecules are by one another, the necessary result is that solvent molecules round the solute molecules are on an average projected towards them, and so rarefied, just as they must be rarefied beneath a free surface. This rarefaction round each solute molecule compensates for the mean decrease in molecular clear space, and increase in internal pressure, which would otherwise occur. Hence the mean internal pressure remains the same in spite of the solute molecules.

There must, however, be a limit to the effectiveness of this process, since with increase in the proportion of solute there will at last be too few solvent molecules, and the internal pressure will, if the solute is a liquid miscible with it in all proportions, gradually become that of the original solute, and will then remain so till the point is reached where the whole of the liquid consists of the original solute. It would seem, therefore, that with increasing proportions of solute the latter is at first "held" by the solvent, then there comes an intermediate state, and finally the original solvent is held by the original solute if the latter is a liquid. Where the two liquids are not miscible with one another in all proportions the intermediate state may be non-existent, or only existent at higher temperatures.

It remains to consider the facts connecting temperature with vapour pressure and molar concentration of the saturated vapour. It was discovered in 1821 by Cagniard de la Tour that when certain liquids are heated to a sufficient extent in a glass tube with space left for their vapour, a point is reached where vapour and liquid become indistinguishable, since, presumably, their densities become equal. It has also been known for long that the vapour pressure of a liquid increases very rapidly with rise of temperature, and at a rate altogether out of proportion to the rise of temperature as indicated by a gas thermometer: also that the vapour, when expanded, no longer follows the gas laws. These facts were investigated quantitatively by Waterston, in the hope of obtaining a clue to the connection between the liquid and gaseous state. He found, by analysis of results published by Regnault and others, and from results of his own, obtained by a quite new method, that when the sixth root of the molar concentration of the saturated vapour is plotted against the absolute temperature, the results at different temperatures lie on a straight line.[1] Figure 12, embodying his data for water, illustrates his conclusion.

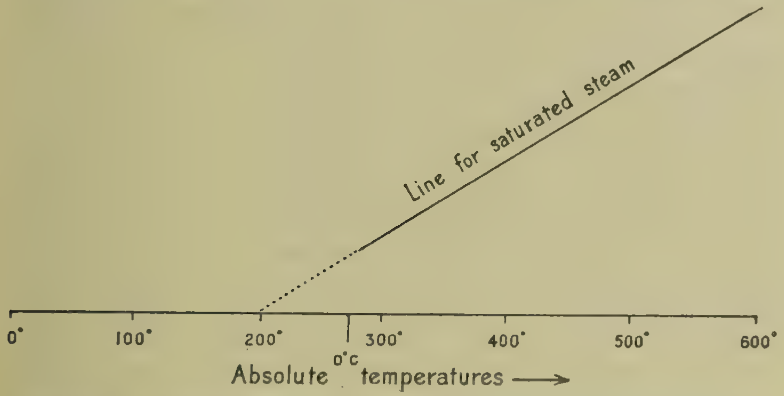

FIG. 12.—To illustrate Waterston's law that when the sixth roots of the densities of saturated vapours at different temperatures are plotted against temperature, the graph is a straight line within wide limits of temperature. The line shown is that for saturated steam, and if it were extended downwards would give a zero of vapour pressure at 194° absolute temperature or $-79°$ C.

---

[1] Waterston, *Collected Papers*, p. 371.

This result, together with the fact that the straight lines obtained did not point towards absolute zero, but to a "zero of vapour pressure," was not intelligible to Waterston; but he saw clearly that he had in his hands a clue to the extension to liquids of the dynamical theory which he had previously applied to perfect gases and to the theory of temperature. In succeeding papers he followed up this clue in his investigations of the relation of temperature to liquid expansion, latent heat of evaporation, and surface tension.[1] Like van der Waals, at a later date, he had discovered a mathematical relationship among phenomena connecting the gaseous with the liquid state; and he was as well aware of the importance of this discovery as van der Waals was of his own discovery.

Very unfortunately, Waterston's paper, which was presented to the Royal Society in 1852, was rejected by the Referees, just as was his previous paper on the dynamical theory of gases and temperature. The rejection of the two papers seems, also, to have prejudiced Waterston so seriously in the eyes of his scientific contemporaries that his further papers on liquids, though published, were disregarded.

In one of these latter papers he also showed that the expansion of a normal liquid with temperature runs parallel with the increase in its vapour density. In other words the value of N diminishes in the liquid parallel with the sixth-power increase in the vapour density. A consequence of this is that if we wish to demonstrate the true critical temperature in a sealed glass tube we must, as was empirically discovered by Natterer, fill the tube to half, and neither less nor more than half, with the liquid if it expands normally.

---

[1] Waterston, *Collected Papers*, pp. 407-528.

As was above shown (p. 246), the concentration of a saturated vapour in contact with its liquid must run down very rapidly as the temperature falls. Can we see any reason why it should run down at the rate corresponding to Waterston's discovery? It was pointed out above (p. 234), that when a molecular aggregate of sufficient size is present in a gas of sufficient concentration at a temperature below the critical temperature, the aggregate is dynamically stable because, though some molecules, in approaching the aggregate, have acquired more kinetic energy than is possessed by the average molecule in the aggregate, this extra energy is immediately spread among the molecules in the aggregate. This particular molecule will therefore not be likely to fly off again; but it will spread its energy among the other molecules, some one of which will be likely to fly off again if the average molecular kinetic energy (or temperature) in the aggregate is already equal to that of the free molecules. Thus the average number of molecules flying off from the aggregate will equal the number flying into it.

But now let us suppose that the temperature is lowered, and is already so far below the critical temperature that the value of πN^2 in the liquid aggregate varies only as N. The lowering will increase, in proportion to the drop in absolute temperature, the surplus of kinetic energy available for averaging out the excess of kinetic energy acquired by the gas molecules as they fly into the aggregate. Hence a larger number of molecules will at first be absorbed than are given off, so that the concentration of the vapour will be diminished. But the fall in temperature diminishes in respect of all six degrees of freedom (three of translation and three of rotation), the kinetic energy of the molecules, while the extra kinetic energy acquired by a molecule as it flies into the aggregate is in respect of only one degree of freedom. The molecules in the liquid and those in the vapour will also, in consequence of the fall in temperature, have lost energy in virtue of the mutual movements of the atoms composing them; but this affects the free and aggregated molecules equally, so does not come into account.

Thus we see that the fall in temperature places six times more energy at the disposal of the liquid for absorbing the gaseous molecules than corresponds to the general fall in molecular kinetic energy. But this at once gives us Waterston's law as to the fall in molecular concentration of the saturated vapour with fall in temperature. We thus appear to have in our hands the explanation of this remarkable law.

In the case of all the substances for which Waterston verified his vapour-density law the molecules contained three or more atoms, and thus possessed six degrees of freedom in taking up kinetic energy. He found, however, that the data did not show agreement with the law in the case of mercury vapour, where, as already mentioned, the molecules only take up energy in respect of three degrees of theoretical freedom.

To judge only from the considerations involved in the dynamical theory of gases, and the extensions of these considerations to liquids as attempted hitherto in this book, we should expect the concentration of the vapour of a substance, whether in the liquid or solid state of aggregation, to fall off continuously according to Waterston's law till absolute zero was reached. It is, however, perfectly clear from Waterston's graphs that this cannot be the case. The straight lines of the graphs for various substances are straight lines so far as he followed them, but point, not towards the absolute zero of a gas thermometer, but to temperatures considerably higher, which Waterston tentatively refers to as zeros of vapour pressure. He had not data for following vapour pressure down to very low pressures, nor were methods then available for doing so. At temperatures below those to which his data applied vapour concentrations doubtless fall off at a slower rate with fall of temperature than corresponds to the extrapolation of his graphs; and this can be seen on his graph for water at temperatures near 0 C. Finally the whole of the considerations on which the dynamical theory depends in its applications to molecular physics cease to correspond with observation. What is perhaps most striking is that the specific heats of all solid substances cease to correspond with Dulong and Petit's law of atomic heats (extended by Kopp to solid compounds). In Waterston's time it was known that the atomic specific heat of carbon and several other solid elements was "abnormal"; but it was not suspected that this "abnormality" was, in fact, the invariable rule for all solid substances at low enough temperatures, and could even be traced in liquids.

Other physical properties, besides specific heats, become equally "abnormal" at very low temperatures, and among them the concentrations of the saturated vapours of the substances are doubtless included. If we tried to use solid carbon in the form of diamond as a thermometer, it would, starting, say, from fixed points at 1000° and 500°, indicate degrees which bore no relation to the degrees indicated by a gas thermometer. At sufficiently low temperatures by the gas thermometer the carbon thermometer would, in fact, cease altogether to indicate falls of temperature shown by the gas thermometer, and would point towards an absolute zero far above that of any gas thermometer. We are accustomed to accept as "objective" the absolute zero to which gas thermometers at ordinary temperatures point; but at low temperatures the very foundations of our ideas of temperature, as originally formulated definitely by Waterston, seem to disappear. It is certainly the same as regards Waterston's law of variation in the concentration of saturated vapours with temperature.

Up to a certain point we can work with complete success on the lines of the Newtonian molecular physics founded by Waterston; but beyond this point the Newtonian mechanical conceptions break down completely. The breakdown became specially evident in connection with the phenomena of radiation. Except in his discussions of biological and psychological subjects Waterston remained to the end of his days a thorough Newtonian; his physical ideas were those of the nineteenth, and not the twentieth century. He came, however, very near to explaining why the sixth root of the density of the saturated vapour of any particular substance varies directly as the temperature, since he suggests that this fact is probably connected with the six directions into which translational movements of molecules may be resolved. At any one time, however, there are only three directions for any particular molecule. He had not taken into account the three degrees of freedom of rotational movement. There are thus, as was pointed out by Willard Gibbs, six actual degrees of theoretical freedom. The curiously simple arithmetical mistake which led Waterston to disregard energy of rotation, and the erroneous experimental data which seemed to justify his mistaken arithmetical calculation, were pointed out by the late Lord Rayleigh in his notes on Waterston's first Royal Society paper.

CHAPTER VIII

THE STEAM-ENGINE AND SOME OTHER HEAT-ENGINES

In Chapter II the theory of heat-engines with gas as the working substance has been discussed. It is now possible to discuss the theory of steam-engines. The steam-engine was the first practical heat-engine, and is still the main source of mechanical power for human purposes, though much of that power is now distributed electrically. The fuel from which the heat for steam-engines is obtained is also usually coal, which is on the whole the cheapest and most widely available source of heat.

A steam-engine differs from any form of gas-engine, owing to the fact that in the cycle of the engine the working substance passes from the liquid to the gaseous state and back. To understand the working of a steam-engine we must bear in mind the nature of this change. Let us, following the treatment of Chapter II, imagine first a schematic form of steam-engine, working on the same principle as Carnot's gas-engine, and which we may call a Carnot steam-engine. Carnot and his successors thought that the gas-engine which he described could theoretically be worked backwards by applying the same amount of power as was gained in the forward working. We saw in Chapter II that he had unconsciously assumed that in the working of this engine it was possible for the cylinder walls and piston to be impervious to heat except in so far as heat was allowed to pass in from the source, and out into the condenser. We may, in the first instance, make the same assumption for the Carnot steam-engine, and consider how such an engine would work, just as we considered this for the Carnot gas-engine.

Let us, then, consider the working of such an engine, with water and steam as working substances, but otherwise similar to the engine considered in Chapter II, and with the top of the cylinder closed. Above the piston there is a perfect vacuum: below it, in the position corresponding to *c d* of Carnot's figure (Fig. 1, p. 33 above, or A of Figs. 2 and 3, pp. 35 and 58 above) is aqueous vapour (very dilute steam) together with a small amount of water condensed in the previous cycle. The temperature at the point A would, if the engine were travelling very slowly, be practically that of the cold body used as condenser, which temperature we may assume to be 0° C. or 273° absolute. The pressure of the steam or aqueous vapour would then be 4.6 mm. of mercury, or 0.006 of an atmosphere, but its volume would be very large indeed, making it very difficult to represent the cycle graphically.

Let us now consider what would happen during compression from A to B, bearing in mind the initial assumption (which, in the case of the steam-engine, we can at once see to be unrealisable) that no heat whatever is lost to the walls by the working substance in this compression. Though the heat produced by the compression could not escape, it would communicate itself to the small amount of condensed water below the piston, the result being that while the water would absorb some of the heat, and to this extent diminish the temperature and consequently the pressure of the aqueous vapour, the pressure of the aqueous vapour would, on the other hand, be increased in consequence of the rise in temperature of the water. This increase would far outbalance the diminution due directly to absorption of the heat of the vapour by the liquid, since the increase in molar concentration of the steam would, by Waterston's law, be proportional to the sixth power of $T - 19.4$, and its pressure would further increase in proportion to $\frac{T}{T_1}$.

In the previous chapter we have seen that the reason for this is that in the liberation of molecules from the liquid state only one degree of freedom of movement is involved, whereas in the rise in temperature of, at any rate, a triatomic molecule of water, six degrees of freedom of movement are involved; so that, owing to the liberation of molecules, the pressure of the saturated vapour rises out of all proportion to rise in absolute temperature. The curve of rise in pressure with adiabatic compression would thus be immensely steeper in the case of the mixture of vapour and small quantity of water than in the case of an equal number of molecules in the gaseous state, starting from the same pressure.

In the Carnot steam-engine under consideration the assumed adiabatic compression is continued until the water and steam, including the steam liberated during compression, have reached 100° C., which is the temperature of the source of heat. The pressure is then 760 mm., and we must also assume that part of the water is still unevaporated. The energy communicated to the water and steam has been converted partly into heat, but partly also into potential energy, owing to partial and complete separation of water molecules.

The source of heat is now applied to the bottom of the cylinder, which is again made pervious to heat. The working substance is allowed to expand, doing work on the piston, but so slowly that the pressure of the steam is not reduced appreciably, since steam is evaporated fast enough to prevent more than very slight fall of pressure, though in virtue of the work being done the steam is losing heat, and condensing momentarily on the piston. The work done during the isothermal expansion is increased in proportion to the extra volume of steam formed. When all the water has evaporated the source of heat is removed. The heat which disappears during the expansion is not merely converted into mechanical energy, but a far larger proportion (at 100° C.) is converted into potential energy in separating the molecules of the water which evaporates.

The steam is now allowed to expand adiabatically, still doing work on the piston, till the temperature reaches that of the condenser. During this expansion there is continuous condensation, since the steam is saturated from the outset. The molecular energy converted into mechanical energy represents not only the kinetic energy present as heat at over 0° C. in the whole of the working substance, but also the potential energy converted into kinetic energy in proportion as the steam condenses and cools.

In the last stage the working substance is compressed until the starting-point, corresponding to A, is again reached. During this stage condensation is occurring continuously, the heat liberated during condensation, as well as heat corresponding to the work done on the steam by the piston, being absorbed by the condenser, and the condensation being so slow that the temperature of the water at the bottom of the cylinder does not rise appreciably above the temperature of the condenser. The pressure during the condensation is thus practically no higher than 4.6 mm.

The energy communicated by the piston to the steam and water during adiabatic compression represents the energy required to raise the whole of the working substance from 0° C. to 100° C., including the energy absorbed as potential energy in separating molecules partially, or completely into steam, during the compression. Similarly the energy communicated to the piston by the steam during adiabatic expansion represents the total energy lost in cooling the whole of the working substance from 100° C. to 0° C., including potential energy lost by the steam which condenses and cools. The kinetic energy absorbed in separating molecules of water during the adiabatic compression is of course the same as what would re-appear owing to the cooling if the process were simply allowed to reverse itself. But when, before the adiabatic expansion, the working substance has been first allowed, as in the assumed cycle, to expand isothermally at constant pressure, the amount of potential energy which is converted into mechanical energy during adiabatic expansion to the temperature of the condenser is increased in proportion to the increased volume of steam which is expanding.

Thus the work done in adiabatic expansion is more than that done in adiabatic compression, and the increase is at the expense of part of the extra potential energy communicated to the working substance during isothermal expansion. The rest of this extra potential energy passes into the condenser as heat of condensation during isothermal compression. In proportion to the condensation in adiabatic expansion the total heat lost in isothermal compression is diminished, and that converted into work in adiabatic expansion increased, thus increasing the net work of the cycle, in return for expenditure of heat in merely evaporating the water.

The higher the temperature of evaporation the greater is the proportion of the extra steam condensed in adiabatic expansion, and the less the proportion to be compressed and condensed in isothermal compression. The less also is the proportion of applied heat converted into potential energy in isothermal expansion, and the greater the proportion converted into mechanical energy, so that with increasing drop in temperature the efficiency will approach nearer and nearer to 100 per cent., or would be 100 per cent. if the temperature to which expansion was carried were absolute zero. It is evident that the efficiency with which the heat applied in isothermal expansion would be converted into mechanical energy would depend on the relative drop in absolute temperature; for the proportion wasted in the condenser diminishes in exact proportion to rise in the absolute temperature of isothermal expansion above that of isothermal compression. At temperatures above the critical temperature isothermal expansion would not be at constant pressure.

If we increased the piston velocity by diminishing the external resistance, the temperature of the evaporating surface, and consequently of the expanding steam, would fall owing to the more rapid evaporation; and similarly the heat liberated at the condensing surface, and to some extent that of the condensing steam, would rise, till at the velocity V the efficiency would fall to zero, just as with a reversible Carnot gas-engine.

It thus appears that transition from the liquid to the gaseous state and back in the cycle of a reversible heat-engine, if such a thing could exist, would not by itself alter the thermal efficiency of a heat-engine working between two temperatures. It is also easy to show that, as Carnot himself pointed out, it would not by itself matter whether the working substance was gaseous, solid, or liquid. It appears that when the action of a heat-engine depends on the expansion of the working substance by the action of heat, its efficiency, if it is reversible, must depend on the relative absolute temperatures of expansion and compression. If for "temperatures of source and condenser" we substitute "temperatures of isothermal expansion and compression," Kelvin's conclusion that the theoretical efficiency varies as $\frac{T - T_1}{T}$ still holds good for a reversible Carnot steam-engine.

Nevertheless it does not follow that in some quite different form of heat-engine, say an osmotic engine in which an osmotic deficiency, and not a mere expansion, was accompanied by the application of heat, the efficiency would, if the engine was reversible, be proportional to $\frac{T - T_1}{T}$. The maximum osmotic deficiency might easily coincide with little or no rise of temperature, and there might be no quantitative correspondence between relative rise of absolute temperature and thermal efficiency.

In the case of a steam-engine it is much more evident than in the case of a gas-engine that a reversible cycle is absolutely impossible. If, with an actual Carnot steam-engine, we continued the compression after removing the condenser, the result would be, if the walls of the cylinder were cold, that the aqueous vapour beneath the cylinder would simply condense on continuing the compression after removal of the condenser. On applying the hot body water would be evaporated from the bottom and would promptly condense on the walls, so that hardly any rise of pressure would occur, though the walls would be slightly warmed, and this would cause a little rise of pressure. We may suppose that the engine had been pushed round a few times, so that the cylinder could become warmed up. The walls would then, if leakage of heat to the outside were practically prevented, have assumed a mean temperature, which would, if there were no leakage to the outside, be midway between the temperatures of the steam during expansion and compression. Just as in the case of the Carnot gas-engine, however, the temperature of the evaporating surface, and, as a consequence, of the steam during expansion, would, owing to the preponderating influence of condensation on the walls in cooling the water by hastened evaporation, be far lower than that of the source, while the temperature of compression would, owing to the steam evaporated continuously from the walls, be above the original temperature of the condenser. The cycle would also be immensely shortened.

During the stage of compression in absence of the condenser the steam would (neglecting the fact that the bottom of the cylinder would still be very cold) be heated and compressed till it began to condense on the walls and finally reached the temperature it would have during isothermal expansion. On applying the hot body steam would begin to be evaporated from the bottom of the cylinder and condense continuously on the walls, its pressure being regulated by the temperature of the surface of the walls and the rate at which it was doing work and simultaneously condensing on the piston. Its temperature would be that of the evaporating surface, which, owing to the rapid evaporation, would be far below that of the source, just as the temperature of water boiling at a low pressure is far below that of the flame heating it. On removing the hot body the steam would cool itself by expansion, and at first by further contact with the walls; but in the later part of the expansion it would be taking up heat and steam from them.

During isothermal compression steam would be continuously evaporating from the walls, thus tending to cool them and raising the temperature of the condensing surface and therefore of the steam blowing continuously into the condenser. The corresponding amount of heat, together with the heat due to the action of the piston on the steam, would pass into the condenser. The temperature of the condenser surface would determine the pressure of this steam, and since the rate of heat-communication by the piston would vary as the square of the piston-velocity, the temperature of the surface of the cylinder bottom, and consequently the pressure of the steam, would be determined partly by the temperature of the cold body and the walls, and partly by the square of the piston-velocity, just as in the case of the temperature and pressure of expansion.

Since the difference in temperature, $T - T_1$, between the steam during isothermal expansion and compression would only be, at best, a fraction of the difference in temperature between the source and condenser, and since the thermal efficiency would evidently depend, just as in the Carnot gas-engine, on the relation of the piston-velocity to the critical piston-velocity at which the temperatures of expansion and compression became equal, the maximal thermal efficiency of the actual Carnot steam-engine could only be a small fraction of the efficiency of the ideal reversible Carnot steam-engine, assuming that in both cases friction was inappreciable.

Whether, owing to the limited piston-velocity and excessive loss of energy by friction, either engine could be so made as to be capable, without being pushed, of completing a cycle at all, is, however, a very doubtful question.

On p. 38 of his book, *Réflexions sur la puissance motrice du feu*, Carnot expresses, in italics, the following conclusion: *The motive power of heat is independent of the agents employed in realising it: its amount is determined solely by the temperatures of the bodies between which transport of heat finally occurs.*" If this conclusion were correct the maximum efficiency of a heat-engine working between two given temperatures would not depend on a certain velocity of the cycle, and could be no greater whatever the mode of working.

When, however, we examine the reasoning which led Carnot to this conclusion we find that his logic was at fault as well as his physics. It is true that, for the reason specified on p. 32 above, any particular heat-engine could not work more efficiently than if it worked reversibly; but it does not follow logically that even a reversible heat-engine of a different character could not work more efficiently between the same temperatures.

Let us now consider what could happen if one reversible heat-engine—say an osmotic engine—worked more efficiently between two temperatures than another—say a gas-engine. The reversible osmotic-engine could, with a smaller consumption of heat, be set to work the reversible gas-engine backwards. The gas-engine would then deliver at the higher temperature more heat than the osmotic engine had taken from that temperature, so that heat would be progressively transferred from the low to the high temperature. The surplus heat could be utilised to drive another osmotic engine forward, and so on continuously so long as heat at the lower temperature was available. We could thus, for instance, employ the heat of the sea to drive motor ships across it. In this there would be no contradiction of the law of conservation of energy, whereas any particular kind of heat-engine which worked more efficiently than when it was working reversibly would involve such a contradiction, as Carnot in effect pointed out.

It was partly a mistake in logic that led Carnot to infer that the method of working can make no difference to the efficiency of different heat-engines working reversibly between the same two temperatures; but the fact that, if their efficiencies were different while yet they were reversible, heat could, by their means, be caused to flow continuously from a lower to a higher temperature without net expenditure of energy, has led independently to the conclusion that Carnot's principle is correct, since heat never does flow from a lower to a higher temperature in this way. The fact that it does not is embodied in what is known as the second law of thermodynamics.

It has already been pointed out, however, that precisely because heat, owing to its nature as chaotic kinetic energy, is always tending to equalise its temperature with that of the environment, a reversible heat-engine is impossible. If we admit the existence of reversible heat-engines in Carnot's sense, we have become involved in the absurdity of heat-engines which work at the expense of heat which is not at a higher temperature than that of the environment. An actual reversed Carnot gas-engine working in a cycle would not take up heat from the cold body during isothermal expansion at the lower temperature except at the expense of extra work done during isothermal compression, so that the heat transferred by it to the hot body would have no other source than the work done on the working substance during compression. There would thus be no net transference of heat from the cold body to the hot body during the work of the coupled engines. Hence neither Carnot's mistaken physical argument, nor the argument from the second law of thermodynamics, has any validity in demonstrating that a Carnot engine working between two temperatures has the maximum efficiency attainable by any heat-engine working between these temperatures.

The argument from the second law of thermodynamics involves the quite wrong assumption, made unwittingly by Carnot and then accepted by Kelvin and the majority of physicists up to the present time, that a reversible heat-engine is possible; and this necessitated the invention of an empirical principle for the purpose of correcting the absurdities to which the wrong assumption leads. The correction was not complete, since it still left open the possibility of a reversible heat-engine. As soon as we see that the assumption in question is wrong we have no further need of any separate "second law" of thermodynamics, since the facts which the "second law" imperfectly embodies follow from the dynamical conception of heat. From that conception it follows at once that heat cannot flow spontaneously from one body to another of *equal* or higher temperature, and must always be flowing from a body of higher temperature to other bodies in contact with it, except in so far as the flow is prevented or reversed by doing of work.

A clear and simple statement of the fallacious argument by means of which it came to be inferred that no heat-engine working between the same two temperatures can have a higher thermal efficiency than a Carnot gas-engine working reversibly will be found in the chapter on Heat-engines in Clerk Maxwell's *Theory of Heat*. From his statement it will be easy to realise the complicated tangle which has arisen out of the acceptance of Carnot's assumption that a reversible heat-engine is possible.

A steam-engine working on the Carnot principle would evidently possess all the practical disadvantages, already pointed out on p. 84, of a Carnot gas-engine. In addition, it would be very wasteful to combine boiler and condenser in a steam-engine. Hence no actual steam-engine has ever worked on the Carnot principle, and all have a separate boiler, and have at present a separate condenser if a special condenser is used at all.

In the reciprocating steam-engine steam is admitted at the end of compression to the space left between the piston and end of the cylinder, care being taken that the rise of pressure in the cylinder is not too sudden. When the piston has travelled a short distance the supply of steam is cut off, and the steam in the cylinder allowed to expand, doing work and therefore cooling itself and condensing. When the expansion has gone as far as found to be profitable it is stopped, and the steam at the same time allowed to escape into the condenser. This does not, or need not, bring the pressure in the cylinder to that in the condenser, since there are fairly narrow passages to be traversed, and steam evaporating from the cylinder walls is all the time blowing into the condenser. The passage of steam to the condenser surface is also more or less hindered by the presence of a little air. During compression this steam, including what is blowing off continuously owing to evaporation from the cylinder walls, is condensed. The condensed water from the condenser is finally pumped back into the boiler.

Supplying heat and potential energy by the free admission of steam is a far more rapid method than by supplying them through the bottom of the cylinder, as in the Carnot engine. Consequently the cooling influence of the walls is comparatively small in relation to the heat supplied, and the available energy per cycle correspondingly greater, just as in the internal combustion engine.

Let us now consider what is happening in the cycle of the engine. We may assume, in the first place, that the steam admitted to the cylinder is saturated. The metal of the cylinder and piston will have been heated up to an approximate mean temperature in previous cycles of the engine, but this temperature will be considerably below that of the steam admitted. Hence the latter will at once begin to condense on the metal, and will also be condensing in proportion as work is done. The temperature and pressure cannot fall, however, so long as steam is being admitted freely enough, and the boiler temperature is kept up, since the place of the condensed steam is at once taken by the fresh steam admitted, and in proportion as the steam tends to cool itself work is instantly done on it, so that its temperature cannot fall. It is primarily in the boiler that cooling occurs if steam is admitted freely enough; and with the engine running in a cycle this cooling has been made up for by heat from the furnace when the next admission of steam to the cylinder occurs. The actual cooling is inappreciable.

Thus the expansion of the steam becomes, for a time, both isothermal and at constant pressure, which pressure may be nearly that of the boiler, but necessarily slightly lower.

It is evident that during the first part of expansion much heat is being absorbed by the metal of the cylinder and piston, besides what is disappearing as mechanical work done. After the steam is cut off the absorption of heat will continue until the temperature of the steam falls to that of the metal. The fall is thereafter due entirely to disappearance of heat in proportion as work is done. The steam cools itself *en masse*, since it does work instantly on the steam which tends to be rarefied by condensation at the piston and cylinder surfaces, and prevents either local rarefaction or difference in temperature at different parts of the steam. The fall in pressure after the steam is cut off is at first faster than if the steam were simply cooling itself by adiabatic expansion, and the condensation is correspondingly faster. As soon, however, as the steam has been cooled to the temperature of the cylinder walls the condensed water begins to evaporate again from the walls, thus making the fall in pressure increasingly slower than if it were adiabatic. If the engine is properly designed for maximum efficiency in its work the expansion is cut short, and steam admitted to the condenser, as soon as the pressure and temperature are reached which are most economical during the compression stroke. To continue the expansion further would be worse than useless, since no further net work would be done, loss by friction would be increased, and the horse-power of the engine would be reduced owing to the waste of time. In an actual engine, however, owing to the effects of friction, as explained in Chapter II (page 78), it is necessary to cut the expansion further short, the heat thus wasted passing into the condenser, in addition to what otherwise passes in during compression; and if increase of horse-power, even at the expense of thermal efficiency, is aimed at, there is still further cutting short, or, what comes to the same thing, the position of the cut-off is altered, so as to permit the entry of more steam per cycle.

The pressure during the compression stroke will correspond to the temperature of the condenser surface, but will necessarily be higher, since the steam has to pass narrow passages on its way to the condenser, and a little air is present in the latter. Not merely the steam which remained uncondensed at the end of expansion, but a considerable additional quantity, evaporated from the hot cylinder walls during compression, will pass to the condenser, and will help in communicating heat to its surface.

If the piston is travelling slowly a great deal of heat, liberated in the condensation of steam, will have time to pass into the metal of the cylinder and piston during the expansion stroke, and thence out again into the condenser (except for what passes to the outside) during the compression stroke. This involves great waste of steam and heat, and with a very slow piston rate and abundant supply of steam per cycle from the boiler nearly all the steam and heat will be wasted in this manner, and in friction, just as the heat is wasted with an internal combustion-engine running too slowly.

The following stages occur in the cycle of an actual steam-engine. In the first stage the condensed water is heated at constant volume of the boiler to the temperature of the boiler. In the second stage steam is allowed to pass into the cylinder as fast as the movement of the piston permits, the steam being provided by evaporation from the boiler. In the third stage the supply of steam is cut off, and the steam in the cylinder allowed to expand till it reaches as nearly as desirable the temperature of the condenser. In the fourth stage the expanded steam is condensed at about the temperature of the condenser surface.

Apart from the purely imaginary reversibility of a Carnot engine the cycle of an actual steam-engine differs radically from that of a Carnot gas-engine; and it is easy to see that Carnot, in imagining the cycle of his engine, was unconsciously transferring to it the characters of a steam-engine. In the steam-engine the steam during isothermal expansion, even with the engine travelling fast, is at very near the temperature of the boiler (regarding this as the source), and at very near the temperature of the condenser surface in compression; whereas nothing remotely approaching to this condition is possible in a Carnot engine, whether the engine is working slow or fast. It was the neglect of heat-leakage on the one hand, and of the limited rate at which heat is conducted on the other, that completely invalidated Carnot's conception of his engine.

In the actual steam-engine heat-leakage exists, just as in the Carnot engine; but the heat which leaks away as steam is condensed on the cylinder walls is replaced through the introduction of more steam from the boiler; and this replacement occurs almost instantly, and not by the slow process of conduction of heat. Thus, unlike the Carnot engine, the actual steam-engine can work quite fast, with a pressure practically corresponding during expansion to that of the steam in the boiler, and during compression to that of steam in contact with the condenser, if sufficient heat is supplied by the fuel to the boiler and the condenser is adequate. The difference in pressure between the steam in isothermal expansion and compression varies, moreover, in accordance with Waterston's law, and not with those of Boyle and Charles. An evident weak point in the steam-engine is the impossibility, owing to friction, with the long expansion which would be needed, of obtaining the whole of the net work which the cycle would theoretically yield ; but this defect becomes, as with the internal combustion engine, of less and less importance as the temperature and piston velocity rise.

With a sufficiently low velocity far more heat is leaking into the cylinder walls in the steam-engine, as in the internal combustion engine, than is converted into mechanical energy during expansion. In both kinds of engine, also, it is only at a certain velocity that, with a given value of $\frac{T - T_1}{T}$, the proportion of net work done reaches a maximum, which cannot exceed $\frac{T - T_1}{2T}$.

As with the internal combustion engine, the value of T is a virtual value, representing the mean temperature at which the work during expansion would be done, assuming the specific heat of the steam to be constant, though it actually increases with the rise of temperature. As, however, most of the work is done at or near the actual maximum temperature, and this corresponds, owing to the increased specific heat, to a higher virtual temperature, the virtual value of T will be near the actual maximum temperature, while with an adequate condenser, the existing value of T_1 will nearly coincide with that of the condenser surface.

As the piston velocity increases, the loss during expansion by leakage into the walls, and by friction, will progressively diminish till, just as with the internal combustion engine, it is balanced by increased loss of heat to the condenser during compression ; and with further increase in velocity the efficiency will begin to diminish and will in addition be reduced by increasing resistance to the passage of steam and friction due to the inertia of reciprocating parts. With an adequate condenser, the increasing loss in compression is due, less to rise in the temperature of compression, than to more and more of the energy communicated by the piston to the steam and water being converted into energy of translation, which is converted into heat at the condenser surface where the translational movement is arrested.

There is, however, an important difference between the steam-engine and internal combustion engine as regards the rate at which condensation can occur ; and this gives the steam-engine a great potential advantage, as well as a greater potential capacity for continuing to work and waste heat with a very low value of $\frac{T - T_1}{2T}$.

When saturated steam condenses on the walls of a cylinder in free communication with a boiler at constant temperature the temperature of the steam next to the walls does not fall, since it is instantly replaced; and any cooling of neighbouring molecules is prevented by their compression at the existing pressure. Thus heat is communicated far faster to any surface by saturated steam than by a gas which obeys Boyle's law on compression. The very rapid communication of heat by saturated steam or even saturated warm air is a familiar fact.

As a consequence of this rapid communication of heat, the rate at which heat leaks into the cylinder walls is far greater than with a gas. The temperature of compression is also far lower. On the other hand, so long as the temperature of the boiler can be kept up, the pressure in the cylinder during expansion does not fall, and with a very slow piston-rate or very low value of $\frac{T - T_1}{T}$ the capacity of the engine for wasting heat without stopping is very great, and only limited by stoppage of the engine owing to increase in frictional loss. Another consequence, however, is that in reaching the maximum thermal efficiency of $\frac{T - T_1}{2T}$, the piston must move far faster, and the horse-power is correspondingly increased. Moreover this rate depends, if admission of steam from the boiler to the piston is sufficiently free, so that the engine is not "wire-drawn," on the capacity of the furnace and boiler for supplying steam. For any given boiler and condenser, however, a point may come where, with further increase in piston-velocity, diminished heat-leakage through the walls is more than balanced by fall in boiler temperature and rise in temperature of the condenser surface, or by frictional loss; and from this point the maximum efficiency commences to decline. T is, however, very nearly the boiler temperature unless serious "wire-drawing" is present, and T_1 the mean temperature of the condenser surface unless serious "wire-drawing" between cylinder and condenser is present. It is only putting the matter somewhat differently if we say that the steam-engine can work much more efficiently and far faster than the gas-engine with the same size of cylinder, value of T, and condenser temperature, owing to the more efficient condensation of steam.

On the other hand, the difference between T and the condenser temperature can be made far greater in the internal combustion engine than in the steam-engine, and in this way an economical internal combustion engine more than makes up for its slow rate of condensation. The virtual temperature of compression has to be far higher in an internal combustion engine than in a steam-engine; but in spite of this the value of $\frac{T - T_1}{T}$[1] can be made higher in certain internal combustion engines than is at present possible in the steam-engine, so that the thermal efficiency is higher in spite of the high value of T_1.

Unless boiler and condenser are properly balanced against one another the temperature of expansion will fall unduly, or else that of compression will rise unduly, in either case unduly reducing the value of $\frac{T - T_1}{T}$; and in proportion as the furnace fails to maintain the temperature of the boiler the value of V will drop, so that the engine begins to work less efficiently, more slowly, and with correspondingly diminished horse-power.

With the steel at present available it would scarcely be practicable, with the help of superheating, to raise the temperature of the steam above about 490° C. or 763° absolute, or to lower the temperature of the condenser surface below 17° C. or 290° absolute. This gives a maximum value for $\frac{T - T_1}{2T}$ of about 31 per cent., or 28 per cent. if 10 per cent. of the heat escapes the boiler; and at present this appears to be the extreme maximum limit of thermal efficiency in steam turbines. On the other hand, the large Diesel internal combustion engine when running at the speed yielding maximal efficiency gives a brake efficiency of 33 per cent. If, however, by multiplying the cylinders we increase the speed of the Diesel internal combustion engine so as to develop greater horse-power for the weight or size of engine, its thermal efficiency falls off rapidly, and the cost of fuel, very high in any case, becomes extremely heavy.

On account of the fairly high piston-velocity which can be reached with even a very low boiler-pressure or value of $\frac{T - T_1}{2T}$, provided that enough steam is available, a steam-engine working very inefficiently will still not stick owing to friction. This fact was very favourable to the early development of the reciprocating steam-engine, but has been almost its undoing in recent times. In the Newcomen "atmospheric" engine the difference between T and T_1 was far less than 50° C., and the boiler-pressure was only that of the atmosphere. Yet a large Newcomen engine could develop great horse-power; and the writer has seen one, about 150 years old, which was still in use for heavy pumping work where abundance of very cheap coal was available, though probably at least half the heat was being wasted by the boiler, and the thermal efficiency could not well have been more than about 1 to 2 per cent. On account of the ease with which great horse-power may be obtained from primitive reciprocating steam-engines with primitive furnaces and boilers, we have become accustomed to regard the reciprocating steam-engine as essentially a very clumsy form of heat-engine, owing its continued use only to local relative cheapness of coal as compared with the liquid or gaseous fuel required for internal combustion engines. The absence of furnace and boiler in the internal combustion engine seems conclusive.

The matter appears very differently, however, when it is considered quantitatively, and in the light of what the reciprocating steam-engine is capable of. In the first place there seems to be no reason why even reciprocating steam-engines should not work normally with a brake thermal efficiency over 20 per cent., provided that the temperatures of the water in the boiler and steam leaving it are kept high enough. By means of water-tube boilers, combined with moderate superheating, we can easily secure the necessary value of $\frac{T - T_1}{2T}$; but in order both to secure this value and to avoid difficulties from deposits of solid matter in the tubes, we must use condensing engines.

The weight and cost of even a compact water-tube boiler and furnace seem at first to weigh down the scales heavily, as against a fast-running internal combustion engine of similar efficiency. On the other hand, however, the steam-engine has important advantages. In the first place the strains to which the cylinder and other machinery are subjected are much less in the steam-engine, so that these parts can be made lighter and are less apt to be costly in upkeep and repairs. In the expansion stroke of an internal combustion engine, particularly if the compression ratio is high, the pressures suddenly developed, especially at starting, are very high. In the cylinder of a steam-engine with the boiler water at a high temperature the pressure is also high, but need not rise appreciably higher, however low may be the piston-velocity. The steam-engine does work in virtue of the difference in pressure between the expansion and compression strokes, and the pressure in compression is very small, while in the internal combustion engine the pressure runs up to a high point in even the compression stroke, so that the total pressure in expansion runs up very high. The cylinder and parts connected with it have thus to be made heavier and stronger with increased frictional losses. As regards weight, therefore, there does not seem to be very much to be gained by using internal combustion oil-engines rather than reciprocating steam-engines of high efficiency, with light water-tube boilers, containing only a small amount of water.

As regards working costs, the steam-engine has of course the great advantage that its fuel, which is normally coal, need cost only about half as much for the same output of work. On the other hand, coal requires more labour in handling and is less convenient to store at sea than oil. But we must not make the mistake of comparing the labour and storage costs of coal with engines of low thermal efficiency with the greatly reduced costs when engines of much higher thermal efficiency are employed. The higher thermal efficiency immensely reduces all these costs. Extra working costs in repairs, replacements, and lubrication with internal combustion engines must also be considered. The higher temperature, as well as pressure, in expansion increases these costs. Altogether it seems hard to justify economically a general use of internal combustion engines for such work as ship propulsion, even if we leave out of account the steam turbine, which can be used with oil as fuel in the exceptional cases where the greater cost of oil-fuel may be justifiable.

When the turbine is used, with water-tube boilers, high-power steam-engines with their appurtenances can be made lighter than internal combustion engines of the same power, and are much cheaper to run. The advantages, with the turbine, of steam and coal for high-power engines on land, and of steam by sea, are generally recognised. The advantages of steam and coal for lower powers of marine engines have recently been pointed out very clearly by Sir John Biles. On the other hand, where cost of fuel does not matter, petrol-driven engines seem to be superior in every way to reciprocating steam-engines.

For small engines the ease with which the fuel can be automatically supplied continuously seems at first to give the internal combustion engine a great advantage. But there appears to be no reason to suppose that coal in the powdered form cannot be supplied in a similar way, securing the advantages of the cheaper fuel. For steam-engines of any kind with a variable load and frequent stops powdered coal seems to present special advantages.

Locomotive steam-engines of all sorts are at present non-condensing engines, and have been since the days of George Stephenson. Let us consider what is gained and what is lost by this arrangement. In the first place we get rid of the condenser; and since no supply of cold water is available for a condenser this may seem necessary at first sight. In the second place we use the exhaust steam for blowing air through the furnace and thus varying the steam supply from the boiler in accordance with the varying requirements of the cylinders. We also use some of the exhaust steam in heating the feed-water to the boiler and injecting this water.

The disadvantages of non-condensing engines are, however, very great, and seem to the writer overwhelming. Perhaps also we can realise best that a condenser is possible when we consider that internal combustion engines, which have for the moment almost cut out steam-engines for road transport, are in part condensing engines, as pointed out in Chapter II. Apart from the nuisance and noise caused by escaping steam and particles of glowing coal shot into the air, such advantages as the non-condensing engine possesses are gained at a very heavy cost in thermal efficiency, with corresponding labour and fuel cost. Even a first-rate modern railway locomotive seems never to reach a thermal efficiency of 10 per cent., and, with its tender, is unduly heavy and clumsy for its power.

With a non-condensing engine the value of T_1 is at least 373°, and in practice always higher. It follows at once that unless we can raise T to a very high figure the thermal efficiency must be very low. We could raise T by superheating, but the slenderness of any advantage so gained as long as the pressure is low, will be pointed out below. To obtain even a moderate efficiency with corresponding increase in horse-power we must raise the temperature and pressure in the boiler. But to do this effectively we require water-tube boilers, and, owing to deposits in the tubes from the continuous supply of new water, we cannot employ effective water-tubes with non-condensing engines. In any case we are heavily handicapped by the high value of T_1.

There seems to be no reason why locomotive steam-engines by road or rail, or even in the air, should not be condensing engines. By using a condenser we can at once go to very high pressures and temperatures in water-tube boilers, at the same time lowering very considerably the temperature of compression and thus reaching a much higher thermal efficiency. We also abolish the noise and nuisance of escaping steam, smoke, and glowing cinders. The furnace can be fed with pulverised coal and air blown in with it, the supply of air and coal being mechanically regulated in accordance with the need for steam. As regards the condenser itself, a water-jacket condenser similar to that of an internal combustion engine would suffice up to moderate sizes of cylinder, or direct air-cooling for very small sizes. This would greatly simplify the condenser and abolish "wire-drawing" in it. The condenser water would be cooled by a "radiator" similar to that used in motor cars, but using evaporation of water to hasten the cooling. The cooling might, of course, be only incompletely effective in reducing the temperature of the condenser to near that of the air; but the effect would be very valuable in increasing the thermal efficiency. Larger and more effective cooling apparatus, aided by evaporation of externally applied water, could, of course, be used with stationary steam-engines where an abundant supply of cold water was not available.

When Watt added the separate condenser to the Newcomen atmospheric engine he effected a great saving of otherwise wasted heat. But he did not see that when he later introduced the principle of allowing the steam to cool itself while doing work in expansion he had at the same time abolished the necessity for a condenser separated from the cylinder. As he well knew, the separate condenser did not prevent leakage of heat through the walls of the cylinder from the steam to the condenser; and whether this leakage is into a separate condenser or a water-jacket is in itself a matter of indifference, provided that the leakage is not too great, whether with a separate condenser or a water-jacket, in relation to the supply of steam per cycle.

The cylinders of steam-engines, and often also of internal combustion engines, are double acting. In a steam-engine this could not imply doubling the work, unless we also doubled the capacities of the boiler and condenser. There is, however, an advantage in starting, and in the reduction of "wire-drawing," assuming that the rate of work is not increased, but the piston-velocity halved. A similar advantage is gained in an internal combustion engine.

A heat-engine may be either designed to work with maximal thermal efficiency, or with maximal horse-power, when it is doing work at its normal rate. When this rate is varied greatly the efficiency or the horse-power must fall off. We can without stopping the engine diminish as much as we like the rate of work done by a steam-engine by diminishing the supply of heat to the boiler or steam to the cylinder, and in this respect the steam-engine has far more "elasticity" than the internal combustion engine; but we must remember that this elasticity is gained at the expense of thermal efficiency. A steam-engine with an excessive load can also without change of gear do the work slowly without stopping, provided that the resistance is not too much for the boiler pressure to overcome; and apart from the safety-valve the boiler temperature goes up as the piston-velocity diminishes. Here again, however, the thermal efficiency becomes low, owing to excessive leakage of heat. With a change of gear the excessive load could be met without diminution in thermal efficiency, since the rate of doing work could be maintained. In any case the engine as a whole, and every part of it, must be designed for a normal rate of work, and if the steam-pressure is low there is no end to the clumsiness and fuel costs of a steam-engine. The clumsiness and fuel costs are, however, not inherent in the steam-engine as compared with the internal combustion engine, but only in the rudimentary design of so many existing steam-engines.

We cannot possibly judge of the thermal efficiency with which the steam is doing its work by the closeness with which the virtual temperatures of expansion and compression correspond with those of the steam supplied from the boiler and of the condenser, and by the efficiency which would correspond to this difference in temperature if leakage of heat could be neglected. In order to measure the efficiency with which the steam itself is actually doing its work we require, not merely a dynamometer record giving the mechanical energy obtained per cycle, but also a calorimeter record giving the heat-energy lost per cycle. If direct heat-loss to the outside were made negligible a calorimeter recording the difference in temperature between inlet and outlet water of the condenser, and volume of water passing, would give the value required. We should then possess the data for quickly judging what thermal efficiencies were being obtained with varying piston-velocities, and simultaneously varying differences in virtual temperature between the steam during expansion and compression. We could also determine at what cost in diminished efficiency maximum horse-power could be obtained. To measure the efficiency of the engine as a whole we should also require, however, a record of the fuel consumed, with its calorific value, and of the heat, and the calorific value of unconsumed fuel, escaping directly to the outside. Unfortunately the indicator diagram of neither a steam-engine nor an internal combustion engine gives a true record of the net work done per cycle by the working substance. The work done by the piston in imparting velocity to the working substance in compression is not represented at all in the diagram. Hence the net work is, if the piston is running fast, a good deal less than the indicated work. If, moreover, we estimate the frictional losses from the large difference between indicated and brake horse-power, we greatly exaggerate the frictional losses. Thus the indicated horse-power or indicated thermal efficiency has little definite significance, and the high values obtained for it with fast-running internal combustion or other engines are apt to be very misleading.

The efficiency of a steam-engine depends upon the completeness with which the heat from combustion of the fuel is communicated to the water and steam in the boiler. If, as happens in older types of boilers, this heat is allowed to leak away freely in various directions, the efficiency of the engine is correspondingly reduced. Even in what, until not long ago, were regarded as good types of boiler, as much as half the heat from the fuel is allowed to escape, so that although the rest of the engine were working in an ideally perfect manner, and there was no loss from friction, the thermal efficiency would be reduced by 50 per cent. With an efficient boiler and furnace, embodying regenerative principles, the leakage of heat need not, in actual practice, reach 10 per cent. of the net calorific value of the fuel when the engine is doing its designed amount of work, and could doubtless be further reduced.

Let us now consider what happens when the steam, instead of being saturated as it enters the cylinder, has been superheated at the boiler pressure. If the steam, as is practically the case with saturated steam up to about 100° C., follows Charles's law in its expansion on superheating, its volume will be increased in proportion to the increase in its absolute temperature. As, however, its temperature is higher, the theoretical maximum efficiency of the engine will be increased in proportion to the increase in $\frac{T - T_1}{2T}$.

But the pressure is no higher, and to realise the large theoretical increase in the work done per cycle the steam, which is unsaturated, will have to expand far further than before. This is clearly impossible unless the stroke of the engine is lengthened to correspond with the superheating, or the entry of steam is cut off earlier, with corresponding diminution in horse-power.

As with great superheating the superheated steam will be very far from being saturated, little or no steam condenses on the walls during expansion, or is evaporated from them during compression. This would greatly diminish the leakage of heat through the walls to the condenser if the difference of temperature between the steam during expansion and compression were the same as without superheating. But as the difference in temperature is increased in proportion to the superheating there may be no actual advantage in respect of diminished leakage.

The more steam is expanded by superheat, the smaller the mass of steam present in the cylinder. The more, therefore, does loss of heat reduce the temperature. To keep this loss down we must run the engine faster. But when we run it faster the engine becomes wire-drawn, since steam cannot enter fast enough, and other frictional losses also increase unduly. Thus the engine works inefficiently, just like an internal combustion engine with an inadequate compression ratio. In an internal combustion engine with inadequate compression we cannot sufficiently increase the piston velocity because the rate of combustion would then be too slow; and with the steam-engine working at inadequate boiler-pressure, we cannot sufficiently increase the velocity because the steam enters too slowly. Excessive superheating may thus apparently result in a loss of thermal efficiency, and it is only with moderate superheating that much gain results.

It thus appears that though great superheating with low steam pressure causes some increase in the work done by the steam, it by no means follows that there is any increase in thermal efficiency. There may, in fact, be a diminution, though a diminution is far less liable to occur with the steam turbine, which much reduces frictional losses. The turbine must, however, be larger, to permit increased expansion. By applying heat which would be used in great superheating to raising the temperature of the boiler we can, however, make sure of an increased thermal efficiency; and the lower the boiler temperature the greater the proportional increase. But where the boiler will not stand a high pressure great superheating will at least give some increase in horse-power. With superheating the amount of steam used per unit of work done is of course always less.

With steam at high pressure, heat absorbed by it in the superheating, and the expansion produced, increase, since the steam becomes more and more full of momentarily aggregated molecules, which are dissociated by the superheating. At the critical temperature half of the molecules are at any moment aggregated together, and isothermal expansion at higher pressure is only possible by superheating.

Potential energy absorbed in superheating thus becomes available in work of expansion, just as if water were still present in the liquid condition within the boiler.

In order, therefore, to get the full benefit of high temperature we must go to very high steam-pressures. We cannot avoid the necessity for high pressures by mere superheating at lower pressures.

With existing reciprocating steam-engines the thermal efficiency is usually very low, even with condensing engines. A brake efficiency less than 10 per cent. is common, and higher than 12 per cent. exceptional. Since internal combustion engines can give a far higher efficiency, and the oil used in them is easier to handle, cleaner, and in the case of heavy oil can be stored and conveyed more conveniently at sea, it is becoming more and more common to substitute oil-engines for steam engines even in very heavy kinds of work such as that of ship propulsion, in spite of the fact that oil commonly costs about three times as much as coal for the same heating power. In the absence of improvement in the steam-engine this substitution would be fully justified.

The low efficiency and power of steam-engines has, most unfortunately for their proper development, and to the serious detriment of the coal industry, been regarded by very many persons as a necessary consequence of the comparatively small rise in absolute temperature which occurs in the cycle of a steam-engine. As we have seen above, this is certainly not the main cause of the low practical efficiency. The main cause is that so many existing steam-engines are antiquated in design, and correspondingly slow in action, clumsy, and wasteful of fuel and labour.

More than half of the heat energy applied in driving a steam-engine or any other form of heat-engine is inevitably thrown away, either in the condenser or directly to the outside. It is only as regards the remaining part that economy is possible; and to secure that economy we must understand what is happening in the engine, and not be misled by Carnot's ingenious conception of a reversible engine, or by the physical mistake in his reasoning. There seems to be no reason why the waste of heat energy in even reciprocating steam-engines should not be reduced to below 80 per cent., and the fact that in reciprocating internal combustion engines the waste has actually been reduced considerably further justifies this estimate. In the internal combustion engine there is just the same possibility of waste, whatever the rise in temperature may be, as in the steam-engine, and it is only by experiment that the internal combustion engine has been so greatly improved. There is every reason for concluding that the improvement of the steam-engine offers a rich field for further research and invention, and that when these are guided by a correct understanding of the steam-engine very great improvements will result. The steam-engine and coal will then regain an established place as sources of power, not merely, as at present, for very powerful engines, but also for road and rail locomotion, and many other purposes. But for the genius of Sir Charles Parsons, the present position of coal and steam as a source of power would be a very dubious one.

With steam at low pressure, however effective we may make the supply and corresponding condensation of the steam, and even if we cut down the frictional losses by using the turbine, it is impossible to obtain a satisfactory output of work from an engine of reasonable size. It is thus in high-pressure steam (not in mere high-temperature steam) that the future of the steam-engine lies, just as it is in high-pressure internal combustion engines (not in mere high-temperature engines) that success in the internal combustion engine has already been realised to so great an extent. For boilers which will not stand sufficiently high pressures there is no future, and water-tube boilers must, it seems, completely replace the old boilers with their greatly limited pressures. Danger from the high pressure is completely neutralised by the greater strength of the water-tubes, and the fact that so small a quantity of dangerously hot water is required to fill them.

When high-pressure water-tube boilers, properly proportioned condensing engines, mechanical firing, with coal-dust where necessary, and corresponding mechanical air-supply on regenerative principles are adopted, the excessive waste of heat and labour with steam-engines, and their excessive size, will be avoided. Until the waste of heat, noise or nuisance of escaping steam and heat, dirt, and human drudgery, so largely associated with the present ordinary steam-engine are abolished, the steam-engine, and the use of coal for driving it, cannot be expected to play the full part which they are so well fitted to play in the production of power. The coal industry of this country will also continue to be unduly handicapped by the cost of distribution of coal and by the preferential use of imported oil for all sorts of purposes to which coal can better be applied. The prosperity of the coal-mining industry is just as dependent now, as it was a century ago, on the effective development of the fast-running small steam-engine.

Several disadvantages in reciprocal heat-engines are due to the violent alternations in pressure, and intermittent flow, during the cycle of a reciprocating engine. These disadvantages are mitigated by using several cylinders, as in the triple expansion engine, but completely abolished by the steam turbine, introduced and developed by Sir Charles Parsons. The striking advantages of the turbine are due largely to the fact that with it the action of the steam is continuous, thus avoiding various fluctuating strains, greatly diminishing frictional loss, and much increasing the horse-power for the size of engine. The steam has also a very free passage into the engine, and from it to the condenser. As regards size, weight, and cost of fuel, the turbine is much ahead of the present internal combustion oil-engine, but so far has only been applied generally to large engines.

The turbine, just like the reciprocal engine, acts in virtue of the bombardment of the movable blades by molecules of steam or gas. The steam is thus cooled, and condenses from the outset if it was saturated. But heat escapes at the same time through the metal without doing work, and by warming the steam below each ring of openings, and re-evaporating its condensed moisture, diminishes the work which would otherwise be done. If, therefore, the turbine is running very slowly, the leakage of heat through the metal will be very great in proportion to the work done, and an immense proportion of hot low-pressure steam will pass continuously into the condenser. With faster running this loss is progressively diminished, but with a certain velocity the efficiency ceases to increase, since the steam and water are issuing with so much velocity towards the condenser that the diminishing loss by leakage is balanced by the increasing loss from the translational energy communicated to the condenser. It is thus easy to see that with the turbine, just as with the reciprocating engine, the maximum attainable efficiency will be only half that which could be attained with very slow running if there were no leakage of heat. With fast running frictional loss is probably negligible, while with any form of reciprocating engine it is considerable, owing to losses by wiredrawing and friction due to inertia of heavy reciprocating parts.

The relative advantages of the turbine as regards thermal efficiency are most evident with low-pressure saturated steam. This is evidently due to the fact that with low-pressure steam friction and the difficulty of getting the steam in and out of the cylinder sufficiently freely tell very heavily, while friction in relation to work done is reduced to a minimum with the turbine, and steam passes in and out very freely. In proportion to the energy contained in it, high-pressure steam can pass in and out more freely, just as a given mass of any other gas passes more freely at high than at low pressure.

With the turbine the frequency of revolution may of course be much greater than the frequency of cycle. This is extremely convenient in the generation of electrical power, but for various other purposes the turbine has to be geared down, and involves this and other somewhat troublesome complications. In spite of them, however, the advantages of the steam turbine for such purposes as ship propulsion are very great. If the circumstances justify the expense of oil as fuel, it can be used most effectively in raising steam for turbines, since the weight and cost of engine is much less.

We can of course employ in heat-engines other substances besides water which pass from one state to another in the cycle of the engine. We might, for instance, employ alcohol or ether or ammonia for an engine working like a steam-engine; but a glance at Waterston's graphs of the variation of the vapour densities of these substances with temperature shows us that for the same difference in temperature they would give a much lower power for their mass than water within the limits of temperature ordinarily available. No working substance seems to be so convenient, safe, and cheap as water. Its high molecular latent heat of evaporation and convenient boiling-point and critical temperature make water a peculiarly suitable agent for employment in a heat-engine. The high latent heat is a measure of the great rise of pressure as water is heated, and it is upon this great rise of pressure that the inherent advantages of the steam-engine depend.

We can look forward to a time in the near future when heat-engines will be properly understood as molecular mechanisms capable of intelligent and easy control. At present the steam-engine is even more imperfectly understood than the internal combustion engine, and correspondingly limited in its usefulness. The empty mathematical formulæ which have been substituted for formulæ applying to the actual world have greatly hindered the development of the steam-engine. It seems that the steam-engine of the future will be comparatively noiseless, clean, of very great power for its size, and responsive at every point to skilled human control without mechanical drudgery. Its output of work will be cheaper in fuel than that of an internal combustion engine of equal power, and will be simply proportional to the heat supplied to it, whatever the nature or cost of the fuel. Hence the cheapest fuel, which is for heavy work coal in the solid form, suitably prepared (not natural oil or oil prepared at much cost from coal), will be used to drive steam-engines and generate mechanically or electrically distributed power wherever cost of fuel is important. The idea that coal will soon be displaced by oil, or that it will become necessary to convert coal on a large scale into oil for the purpose of obtaining power, is devoid of scientific foundation.

In Chapter IV at p. 171, reference has been made to the probability that the transference of water which seems to occur within each segment of a muscle fibril, so that each segment becomes flatter and wider, is an osmotic phenomenon. Clearly, however, there must be some source for the mechanical energy, and the known facts with regard to the respiratory exchange and oxidation of organic material in active muscle point to oxidation of organic material as the source.

It therefore seems probable that when each segment of a muscle fibril is excited easily dissociable molecules are dissociated in one part of each segment, and absorb heat energy furnished by oxidation caused by the excitation, the osmotic deficiency producing, on the small scale existing, an instant transfer of water from the other parts of the segment, and accompanying contraction of the fibril. As soon as the excitation passes off, and the oxidation ceases, the dissociated molecules will recombine, thus restoring the original equilibrium, so that the displaced water returns to its original position and the fibril relaxes, with liberation of heat.

It is known that the thermal efficiency of muscular contraction is, like that of any actual heat engine, considerably less than 50 per cent., though apparently more than 25 per cent. under favourable conditions. It is also known through the investigations of A. V. Hill that of the heat produced in a single muscular twitch only the smaller part appears during or before contraction: the greater part appears during relaxation or recovery. Assuming that the energy of oxidation is mostly absorbed in compensating what would otherwise have been a fall in temperature, we should expect heat to appear in the manner it does as the portions of the molecules recombine. All the evidence seems to point also to the inorganic constituents, and particularly potassium and phosphorus compounds, of muscle playing an essential part in the process of dissociation and recombination. As, however, the dissociation has no definite relation to increase in absolute temperature, the thermal efficiency of the muscle has no relation to the fraction $\frac{T - T_1}{T}$.

On this conception a muscle-fibre consists of chains of innumerable tiny heat-engines, acting in unison when they act normally, and each performing several thousand cycles per minute. Such a conception becomes possible at once when we get rid of the implications of Carnot's assumption that reversible heat-engines are possible.

As in the case of so many other organic activities, we seem to find, in muscular contraction and other spontaneous protoplasmic movements, molecular machinery established and maintained at the beck and call of an "organic co-ordination," which, from the Newtonian mechanistic standpoint of the present book, is essentially unintelligible. We only deceive ourselves if we imagine that either past progress has brought us any nearer to a mechanical explanation of this co-ordination, or that future progress is capable of doing so.

Many ordinary physical phenomena, and particularly those observed at very low temperatures, are equally unintelligible from the Newtonian standpoint. It was not, however, the object of this book to discuss such phenomena or recent developments of physical theory in relation to them, but only to help in carrying the Newtonian interpretation as far as possible in connection with molecular physics and its many-sided applications. In order to realise the limitations of the Newtonian interpretation we have only to ask ourselves such questions as why it is that the specific heats of substances increase or diminish in the particular way they do with temperature, or why a gas with a monatomic molecule has so low a specific heat, or why atoms occur as definite species, and maintain their specific internal activities as they do, just like living organisms.

It is now evident that even in what we distinguish as the inorganic world mechanical interpretations are based on something far deeper. Nevertheless it remains true that these same mechanical interpretations, provided that we do not attempt to extend them to what they can never interpret, are of the utmost practical use.

INDEX OF AUTHORS REFERRED TO

[*For Subjects see Table of Contents*]

A

Aitken, p. 241
Amagat, pp. 13, 252
Andrews, pp. 216, 226
Arrhenius, pp. vi, 163-166, 197-201
Avogadro, p. 3

B

Beckmann, p. 210
Berkeley, Earl of, pp. vi, 139, 145-151, 155, 188, 274-276
Bernoulli, D., p. 5
Berthelot, D., p. 232
Biles, Sir John, p. 313
Black, p. 31
Blagden, p. 175
Boltzmann, p. 10
Boyle, p. 2
Burdon Sanderson, p. 172
Burton, pp. 188, 274-276

C

Cagniard de la Tour, p. 278
Callendar, p. 145
Carnot, pp. ix, 7, 29-85, 102, 286-293, 296-299, 304
Charles, p. 2
Clapeyron, p. 37
Clausius, pp. 10, 39, 106
Clerk, Sir Dugald, p. 93

D

Dalton, pp. 1, 2, 12
Davy, Sir Humphry, p. 31
Descartes, p. 171
De Vries, pp. vi, 133, 134, 162
Diesel, pp. 98, 101
Donnan, p. 168
Dulong, p. 283
Dupré, p. 14
Dutrochet, p. 130

E

Eötvös, p. 261

F

Fleuss, p. 174

G

Gassendi, p. 1
Gay-Lussac, pp. 2, 3
Gerhardt, p. 3
Getman, p. 178
Gibbs, Willard, pp. 10, 285
Graham, pp. 4, 109, 131
Guldberg, p. 222

H

Hartley, E. J. G., pp. 145-151, 155, 158, 274-276
Helmholtz, p. 39
Henry, p. 202
Hill, A. V., p. 329
Hirn, p. 14

J

Jeans, Sir James, p. 19
Jolly, p. 130
Jones, H. C., pp. 177, 178
Joule, pp. 31, 39, 102

K

Kant, p. 217
Kelvin, Lord, pp. ix, 7, 38, 39, 40, 106, 293
Kopp, p. 284
Krönig, p. 9

L

Laplace, p. 257
Loeb, p. 168

M

M'Dougall, p. 172
Mariotte, p. 2
Maxwell, Clerk, pp. 10, 44, 299
Mayer, pp. 20, 39
Mohr, p. 22
Morse, pp. 139, 151-158
Mosotti, p. 220

N

Nägeli, p. 131
Natterer, p. 281
Newton, pp. 105, 106, 284, 330

O

Otto, p. 93

P

Parsons, Sir Charles, pp. 323, 325
Porter, p. 274

R

Ramsay, Sir William, pp. 249, 255, 260, 261
Ramsden, p. 264
Rankine, p. 236
Raoult, pp. vi, 175-215, 268-270
Rayleigh, Late Lord, pp. 5, 10, 28, 285
Rumford, Count, p. 31

S

Shields, pp. 260, 261
Smits, p. 187
Stephenson, George, p. 313
Still, pp. 97, 98

T

Traube, p. 132

V

Van der Waals, pp. vi, vii, ix, 1, 15, 141, 224-226
van't Hoff, pp. v, ix, 1, 24, 112, 119, 135-139, 140-143, 162, 180-184, 190, 196, 208-215

W

Waage, p. 222
Waterston, pp. vi, vii, viii, 4-9, 19, 37, 40, 190, 220, 246, 255, 256, 259, 279-285
Watt, pp. 29, 315
Wilson, C. T. R., p. 241

PRINTED IN GREAT BRITAIN BY OLIVER AND BOYD, EDINBURGH.

Medium 8vo. lxviii + 712 pages

The
Collected Scientific Papers of John James Waterston

EDITED, WITH A BIOGRAPHY, BY

J. S. HALDANE, M.D., LL.D., F.R.S.

Fellow of New College, Oxford

OLIVER AND BOYD
EDINBURGH: TWEEDDALE COURT
LONDON: 33 PATERNOSTER ROW, E.C.
1928

CONTENTS

I. Physiological Psychology.

Thoughts on the Mental Functions, 1843.

II. Gases and the Theory of Heat.

Note on Molecularity, 1843. Note on The Physical Constitution of Gaseous Fluids and a Theory of Heat, 1843. The Physics of Media that are composed of Free and perfectly Elastic Molecules in a State of Motion (with the late Lord Rayleigh's Introduction and Notes, and an abstract of a paper on A General Theory of Gases. An Account of a Mathematical Theory of Gases, 1845. The Deviation from the Primary Laws of Elastic Fluids, 1857. The Theory of Sound, 1858.

III. Liquids and their Vapours.

A General Law of Density in Saturated Vapours, 1851. A General Law of Density in Saturated Vapours. Capillarity and its relation to Latent Heat, 1858. Evidence of a Graduated Difference between the Thermometers of Air and Mercury below 100° C., 1858. A Law of Liquid Expansion, 1861. The Expansion of Water at High Temperatures, 1863. Liquid Expansion, 1864. Certain Thermomolecular Relations of Liquids and their Saturated Vapours, 1868.

IV. Miscellaneous Physical and Chemical Subjects.

A New Dynamico-Chemical Principle, 1831. The Integral of Gravitation, and its consequents with regard to the Measure and Transfer or Communication of Force, 1858. Chemical Notation in conformity with the Dynamical Theory of Heat and Gases, 1863. Chemical Notation, 1863. The Application of the Vapour-Density System of Notation to express Chemical Reactions, 1864. Some Electrical Experiments and Inductions, 1865. Some Electrical Experiments, 1865. The Thermal Action in the Wire of a Voltaic Circuit. Electric Conduction, 1865.

V. Astronomical Subjects.

Finding the Latitude, Time, and Azimuth, 1842. A Graphical Method of predicting Occultations, 1844. Remarks on the Comet, 1845. A Direct Method of determining the Distance of a Comet by Three Observations, 1846. A Graphical Method of computing the Excentric Anomaly, 1850. Thoughts on the Formation of the Tail of a Comet, 1859. Observations on Solar Radiation, 1861. The Change that would take place in the Elements of the Earth's Orbit by a Sudden Accession to the Sun's Mass, 1866. The Change that would be induced in an Elliptic Orbit if the Intensity of the Force of Gravity were influenced by the Centripetal Velocity of the Orbital Body, 1867. The Heat of the Stars, 1878. A Solar Thermometer Couple to measure the Heat of Daylight.